GENEBIO PHENYX

User Manual

Phenyx Web Interface

www.phenyx-ms.com

All intellectual property rights on this User Manual, as well as on the Phenyx software, belong to Geneva Bioinformatics (GeneBio) S.A. No part of this User Manual may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording or any information storage or retrieval system, without permission in writing from Geneva Bioinformatics (GeneBio) S.A.

© 2005 Geneva Bioinformatics (GeneBio) S.A., Avenue de Champel 25, CH-1206 Geneva, Switzerland. All rights reserved.

Phenyx uses Perl software.

© 1989, 1991 Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111, U.S.A.

Phenyx uses OLAV algorithms.

© 2003 Geneva Bioinformatics (GeneBio) S.A., Avenue de Champel 25, CH-1206 Geneva, Switzerland.

Phenyx uses Mascot result files.

© 2003 Matrix Science Ltd., 8 Wyndham Place, London W1H 1PP, United Kingdom.

Phenyx uses SEQUEST result files.

© 2005 Thermo Electron Corporation, 81 Wyman Street, Waltham, MA 02454-9046, U.S.A.

Phenyx uses X!Tandem results files.

© 2004 The Global Proteome Machine Organization

Phenyx uses X!Hunter results files.

© 2005 The Global Proteome Machine Organization

Trademarks:

SEQUEST is a registered trademark of the University of Washington.

Phenyx and GeneBio are trademarks of Geneva Bioinformatics S.A.

Excel, Word and Internet Explorer are trademarks of Microsoft Corporation.

Mozilla Firefox is a trademark of the Mozilla Foundation.

iTRAQ is a trademark of Applera Corporation.

Rather than put a trademark symbol in every occurrence of trademarked names, we state that we are using the names only in an editorial fashion, and to the benefit of the trademark owner, with no intention of infringement of the trademark.

Terms and conditions. All goods and services are sold subject to the terms and conditions of sale of the company that supplies them. A copy of these terms and conditions is available on request.

Limitation of liability. The information in this User Manual is subject to change without notice and should not be construed as a commitment by GeneBio. GeneBio is not responsible for the misinterpretation of results obtained by following the instructions in this manual.

Technical support. GeneBio provides technical support for Phenyx in accordance with the terms of the License Agreement. Please contact us at support@phenyx-ms.com.

Databases. Phenyx provides access to several databases on the Internet. It is the responsibility of the user to acquire the database licenses, if needed.

Contents

Introduction 1

Requirements 1
General workflow 2
Helpful tips 2

Phenyx Desktop page 3

Navigation bar 3
Job report 3
Job menu 6

Submission page 8

Profiles 8
Title 8
Database(s) 8
Search engine 11
Acceptance parameters 18
Peak lists 19
Server log 19

Management Console page 22

My defs 23
Define a new enzyme 24
Edit a user-defined enzyme 26
Delete a user-defined enzyme 27
Define a new modification 29
Edit a user-defined modification 30
Delete a user-defined modification 30
Group defs 32
Account Settings 32
Databanks 32
Jobs management 32
Browse files or export job results 32
Import Phenyx jobs or jobs from other search engines 34
Utilities 35

Status page 36

Parameters page 38

Proteins Overview page..... 39

Best-scoring protein table 39
Protein details..... 40
Peptide match table..... 41

Protein Match Details page 45

Protein information table..... 45
Peptide match table..... 46
Sequence views..... 48

Compounds Overview page 50

Functions 53
Validate peptide matches 54

Peptide Match Details page 55

Mass spectrum 56
Fragment ion tables..... 56

Results Comparison page 59

Insert jobs in comparison table..... 59
Define a new comparison 59
Remove jobs from a comparison table 60
Results comparison table..... 61
Protein detail list 62
Export options 63

Detailed Results Comparison page..... 65

Detailed results comparison table 65
Export options 66

Understanding Phenyx: the scoring system 67

The MS/MS scoring model 67

How a score is computed 68
How a score is normalized 68
How Phenyx retrieves amino acid modifications not specified by the user 69
Reproducibility of z-Scores and p-Values 71
Available scorings 73

Understanding Phenyx: the effect of submission parameters on result accuracy 74

Limit the search space size 74
Change the taxonomy or AC list 76
Change the cleavage mode 76
Change the number of missed cleavages 78
Change the number of variable modifications 79
Search in one round or two 80
One-round usage 80
Two-round usage 81
Two-round search parameters 82
Activate conflict resolution 83
How Phenyx resolves conflicts 83
Some points to consider 84

Introduction

Updated: May 10, 2008

Welcome to Phenyx - a software platform for the identification and characterization of proteins from mass spectrometry data.

This User Manual is dedicated to Phenyx users.

Depending on the Phenyx solution you have chosen,

- Phenyx Public server, for initial and restricted submissions
- PhenyxOnline, for a secured environment without hardware/IT issues
- PhenyxServer, for a highly customized platform, also in large cluster environments

the access to some features may differ. It is then noticed in the present Manual.

Requirements

To get started, there are the following requirements:

- Web browser (Mozilla Firefox or Internet Explorer 6.0)
- Peak list files
- Phenyx username and password

Phenyx Public server users have to create their login credentials from the Login page at phenyx.vital-it.ch/pwi).

PhenyxOnline customers, please contact support@phenyx-ms.com for any login issue.

PhenyxServer customers, please contact your admin to create your login credentials.

General workflow

The usage of the Web site is intuitive and follows a step-by-step process.

1. Log in: Connect to the Phenyx Calculation Unit on the Login page (phenyx.vital-it.ch/pwi) and access the Phenyx Desktop.
2. Submit: Specify the parameters on the Phenyx Submission page.
3. View results: See different summaries of the identified proteins and peptides on the Proteins Overview, Compounds Overview, Protein Match Details, Peptide Match Details or even the Results Comparison pages.
4. Generate reports: Export Excel, XML or text formatted files of your job results.

Helpful tips

There are several practical ways to get the most out of the Phenyx Web Interface:

- Allow cookie handling in your browser.
- Get a paper copy of a page using your browser's Print option.
- Read Phenyx documents that are available at www.phenyx-ms.com/documentation/documentation.html.
- Access the User Documentation page at http://phenyx.vital-it.ch//docs/index.html or click on the Documentation link at the top of the Phenyx Desktop. The manual is available online or as a PDF file. Further knowledge of Phenyx including facts about the Phenyx Calculation Unit is supplied in the FAQ and Troubleshooting sections.
- Contact GeneBio at any time with questions or comments by sending e-mail to support@phenyx-ms.com.

Phenyx Desktop page

Updated: May 10, 2008

The Phenyx logo is included at the top of this home page.

Navigation bar

Logged As: The username of the person currently logged in and the type of account privileges (in parentheses).

Log Out: Quit the Phenyx system.

A set of links to the major sections of the Phenyx Web Interface is given in the header row.

Submission: Click on the Submission link to go to the Submission page where you can specify search parameters and peak lists. Send a submission file to the Phenyx Calculation Unit for processing.

Result Comparison: Compare the results of multiple jobs (including output from SEQUEST-Bioworks, Mascot or X!Tandem). You can also use this functionality to open one job for a quick and easy visualization.

Management Console: Specify user preferences and access the Phenyx functionalities. You can define enzyme cleavage and amino acid modification rules. You can import jobs from other third-party software. You can also transform peak lists from one format to another and produce various reports.

Documentation: Access the online documentation section including the Phenyx user manual.

Job report

A summary of your jobs is given in the center of the page.

Completed: Click on the Completed link to display only jobs that finished successfully in the report.

Running: Click on the Running link to display only jobs that are currently being processed in the report.

Error: Click on the Error link to display only jobs that failed to process in the report.

Pending: Click on the Pending link to display only jobs that are in the queue awaiting processing in the report.

All: Click on the All link to display every job submitted to the Phenyx Calculation Unit.

Actions on Selected: Perform operations on certain jobs. Select a job by clicking on the box at the beginning of its row. If a job is selected, then the box is checked, the row is highlighted in blue and the Job Menu is opened.

- Delete: Remove the selected jobs from the report.
- Kill: Cancel the processing of the selected jobs.
- Compare: Insert the selected job(s) in a table on the Results Comparison page.
- Parameters: See a comparison of the submission parameters for the selected jobs.
- Deselect all: Dismiss every selection.

Refresh: Click on the Refresh button to update the report.

Top: Click on the Top button to move to the first (i.e. most recent) thirty jobs in the report.

Up: Click on the Up button to move backwards through the report. Thirty jobs are displayed at a time.

Down: Click on the Down button to move forwards through the report. Thirty jobs are displayed at a time.

ID: Job identification number. Click on the ID link in order to open the Job Menu.

User: The username of the person currently logged in and any other usernames whose jobs you have been granted permission to look at.

Status: Current status of the job. Click on the Status flag to go to the Status page. If the job has already finished successfully, then you link directly to the Proteins Overview of the results.

- Completed: Jobs that finished successfully are green.
- Running: Jobs that are currently processing are blue.
- Error: Jobs that failed to process are red.
- Pending: Jobs that are in the queue awaiting processing are yellow.

Date: Year-Month-Day of job processing.

Title: The search description specified on the Submission page.

Comment: Encoded description about the processing of a job. When a job completes successfully, then the message tells you the:

- Number of proteins identified (i.e. matches) not including the proteins that are found in the subset
- Number of identified peptides (in parentheses)
- AC numbers followed by the Phenyx score for each protein (in parentheses)

These values are updated upon manual validation. The message [user selected] then appears followed by the number of user validated peptides.

If a job is Running, a comment about the progress is given. If a job did not proceed successfully, a short error message appears. To find out why a job failed and perhaps to attempt to fix the problem yourself, click on the Error Text link to go to the Troubleshooting page.

Job menu

A list of functions that can be performed on the selected job appears on the left side of the page. Click on a job's check box or ID to open the menu.

Resubmit: Opens the Submission page with the parameters set as in the selected job. You can however make changes. The peak lists that were previously used will automatically be uploaded if you keep the Resubmitted Peak Lists box checked. You can also add new peak list(s) with a format different from that of the resubmitted data.

Parameters: See a concise restatement of the main submission parameters including the specified title, taxonomy, scoring model, modifications, thresholds, etc.

Proteins Overview: A comprehensive summary of the identification results is presented.

Compounds Overview: All mass spectra that were matched during the search are listed with their corresponding peptide identifications and protein affiliations.

Export Excel: Exports the results of the job as an Excel (.xls) spreadsheet in either the Microsoft application or a new browser window. Use the Save As option to store a copy of the data to your hard disk.

Export XML: Exports the Phenyx result file in Extensible Markup Language format to a new browser window. Use the Save As option to store a copy of the data to your hard disk.

Export Text: Opens a new browser window with the Identification jobs file browser tool for compiling specified Phenyx results into a text format. These text-formatted reports take into account any manual validation done by the user. Select a job number and template from the drop-down lists. Click on the Report button. The tab-delimited report appears at the bottom of the pane. With Mozilla Firefox, you can save the report (and import it into Excel, for example) by right clicking and choosing This Frame > Save Frame As in the pop-up menu. With

Internet Explorer, you can copy and paste the report into a text editor or Excel by right clicking and choosing Select All in the pop-up menu. Click on the Report Details link for further explanations of the different templates. If you are using a local installation of Phenyx, you can define your own templates and make them available in the list. As a user of a locally installed version, you can change the name of your job in the title.txt file. However, the new name is not updated in the reports (i.e. only the original job title appears).

Note that you can use the Identification jobs file browser tool to extract information about iTRAQ labeling reagents. Select the job ID, click on Export Text in the Job Menu and then choose the template called default.ACPept_itraq. The MS/MS intensities of the reporter ions (m/z 114, 115, 116 and 117) are reported in the table of identified proteins and peptides. Save the report in Excel as described above to analyze and quantify your samples.

User Permissions: Opens a new browser window with the User Permissions tool for defining access to the given job. By default, the current user's account appears with all permissions. Fill in the User/ Group Name field to add new people. Then specify the level of permissions. Read access allows the job to be viewed only. Write access allows the job to be viewed and edited. Delete access allows the job to be viewed, edited and erased by the specified user or network of users. Click on the Save button. The job is now listed on the Phenyx Desktop of the user(s) who received access rights to it. The account name of the person who originally created the file is shown under the User column in the Job Report.

Submission page

Updated: May 10, 2008

The Submission page is divided into several modules.

Profiles

This function enables the user to manage different representative sets of data for submission. The name appearing in the box is the current profile used in the Submission page. To load a different profile, select the corresponding name in the drop-down list. Click on the red arrow to expand the menu options.

Save As Profile: Click on the Save as Profile button to store the current page settings as a new profile. Enter the name of the profile in the popup window. Click the OK button.

Set as Default: Click on the Set as Default button to store the current page settings as the default profile.

Delete Profile: Click on the Delete Profile button to remove the selected profile.

Title

Descriptive name for the data identification query (i.e. job).

Database(s)

Parameters used to define the search space are found here.

Database(s): Available protein and/or nucleotide sequence databases are given in the drop-down list. Information such as the release number and sequence types (amino acid or nucleotide) are displayed [in brackets]. Select the sequence database(s) for the search. Use the Shift or Ctrl keys to make multiple selections. If you choose a highly annotated database such as UniProt_Swiss-Prot, many of the

annotations are considered during the search. For example, Phenyx processes the entries and generates separate entities for each splicing variant or searches for post-translationally modified amino acids if the corresponding annotations exist. Licensed users of Phenyx are able to create their own repository of FASTA sequences with the Private Databank tool in the Management Console. In special cases, other databases can be added to the public server. Please contact GeneBio at support@phenyx-ms.com for further information.

- UniProt: The Universal Protein Resource (UniProt) is the world's most comprehensive catalog of protein sequence and function data that was created by the consolidation of the Swiss-Prot and TrEMBL databases. Go to http://beta.uniprot.org/ to learn more.
- UniProt_SwissProt (named uniprot_sprot in the pwi): Database developed by the Swiss Institute of Bioinformatics (SIB) and the European Bioinformatics Institute (EBI). It is a curated database meaning that the entries (post-translational modifications, splicing, mutations, etc.) are generated with relevant annotations from scientific publications or expert data. Refer to http://www.expasy.org/sprot/ for additional information.
- UniProt_Swiss-Prot, reversed (named uniprot_sprot_rev): Reversed version of uniprot_sprot. All sequences have been reverted, and each AC is followed by a "_rev"
- UniProt_TrEMBL: The computer-annotated supplement to Swiss-Prot. It contains protein translations for all DNA coding regions recorded in the European Molecular Biology Laboratory (EMBL) database as well as additional protein sequences extracted from the literature or submitted to UniProt which are not yet integrated into Swiss-Prot. TrEMBL is part of the International Sequence Database Collaboration between the DNA Data Bank of Japan (DDBJ), GenBank from the National Center for Biotechnology Information (NCBI) and the EMBL database. Refer to http://www.expasy.org/sprot/ for further details.
- UniProtKB_sptr: A concatenated form of uniprot_sprot and uniprot_trembl
- IPI: The International Protein Index (IPI) is administrated by the EBI. This database cross-references the human, rat and mouse proteomes found in Swiss-Prot, TrEMBL, RefSeq and Ensembl. Go to http://www.ebi.ac.uk/IPI/IPIhelp.html for more information.

- NCBInr: The collected, non-redundant set of data from a global Entrez query. Entrez is the Life Sciences Search Engine created by the National Center for Biotechnology Information (NCBI). Go to http://www.ncbi.nlm.nih.gov/gquery/gquery.fcgi for more information.

Taxonomy: Classify a kingdom, phylum, class, order or species. This significantly narrows the search criteria. A degree of certainty about the origin of the protein(s) is needed. The main groups are given below. By choosing one of the main groups, all sub-taxons are automatically selected.

- Root: Searches all taxons.
- Bacteria
- Eukaryota
- Archaea
- Viruses
- Other root: Searches all other taxons except Bacteria, Eukaryota, Archaea and Viruses.
- NO_TAXONOMY: Only searches accession numbers input by the user in the AC List.

Licensed users of Phenyx have the possibility to set their own taxonomy tree as defined by the NCBI Taxonomy center (http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy), please contact use at support@phenyx-ms.com if you need help, or if you want to have a trial setting on the Public server.

AC List: Specify a list of accession numbers to search in the specified database(s). More than one AC number can be entered by using a comma (,) to separate them. If you input AC numbers, then an OR operator is used to consider the Taxonomy and the AC List (i.e. both the specified Taxonomy and AC numbers are searched). If you wish to restrict your search to the list of AC numbers only, then select NO_TAXONOMY in the Taxonomy drop-down list.

Search engine

Instrument Type: Specify the type of mass spectrometer that was used to collect the data. The algorithm that generates the score between experimental and theoretical masses is based on several instrumental properties such as the types and charge states of the fragment ions produced. If the exact instrument that was used to collect the MS/MS data is not listed then opt for a similar method.

Scoring Model: Determine the detailed scoring scheme to be applied to resolve your data. This parameter is dependent upon the instrument type. Note that the mass tolerance for the fragment ions is included in the algorithm and cannot be set separately.

Note: information into the Scoring Model top-down menu refers to the fragmentation process. Where the parent mass is detected is not considered. For instance, the scoring model CID_LTQ_scan_Orbitrap_6ppm refers to data that have been fragmented in LTQ and detected in Orbitrap with a fragment error tolerance as up as 6ppm.

Default Parent Charge: The charge of the protonated ($[M + nH]^{n+}$) or deprotonated ($[M - nH]^{n-}$) parent ion as detected in the formatted peak list file by Phenyx. The peak picking software that comes with your instrument assigns charges to the parent and fragment ions based on the observed mass-to-charge ratios. These values are generally included in the peak lists. If two or more integers are separated by commas, then all charge states listed are computed.

- 1: One charge (also commonly referred to as singly-charged or the +/- 1 charge state).
- 2: Two charges (also commonly referred to as doubly-charged or the +/- 2 charge state).
- 3: Three charges (also commonly referred to as triply-charged or the +/- 3 charge state).
- 4: Four charges (also commonly referred to as the +/- 4 charge state).

Trust Parent Charge: The peak picking software that comes with your instrument assigns charges to the parent and fragment ions based on

the observed mass-to-charge ratios. The file format generally includes this information. Do you trust the peak picking software and the file format?

- Yes: The supplied charges are most likely correct. Only the assigned charges should be calculated by Phenyx.
- Medium: There is some uncertainty about the assigned charges and therefore all combinations should be taken into account by Phenyx. For example, a precursor ion charged 1+ is considered as one charge. Two charges are considered to be 2+ or 3+. This increases the calculation time.

Number of Rounds: One or two sets of calculations can be carried out on the data. A second round of calculations is used to narrow or fine-tune the results obtained from the first round of scoring because only the accession numbers that fulfilled the first round criteria are processed during the second round. In other words, the first round searches the whole database(s), whereas the second round only looks at the proteins with satisfactory scores to have passed the first round.

By default only the Round # 1 box is checked and the related search parameters are visible (i.e. the red arrow is pointing down). If you choose to do two rounds of calculations, then check the box corresponding to Round # 2. The same parameters specified in Round # 1 will appear under Round # 2. Click on the red arrow to collapse the search parameters (i.e. the red arrow is pointing right). Find an example of possible settings for a two-round search below. To learn more about the reasoning behind these selections go to Understanding Phenyx: the effect of submission parameters on result accuracy at the end of this manual.

Parameter	**Round 1**	**Round 2**
Fixed modification(s)	Yes	Yes
Variable modification(s)	None	Yes
Missed Cleavage(s)	1	3
Cleavage Mode	Normal	Half cleaved

Parameter	Round 1	Round 2
Conflict Resolution	None	Yes
Peptide z-Score	6.0	5.0
Peptide p-Value	1.0e-6	1.0e-5
Turbo	Yes	None

AA Modif.: Amino acid modifications, select any chemical modifications, post-translational modifications (PTMs), cross-linking, etc. from the drop-down list. The description of every listed amino acid modification is available in the Management Console. Click on the **Def. (Definitions)** link to go to the Management Console page. Click on Residue Modifications to view or edit the definitions. Modifications can be specific to the user (My defs) or common to all users (Group defs).

Use the Shift or Ctrl keys to make multiple selections. If a modification is selected, then it appears in the **AA Modif. Details** table to the right. Specify the Type and Tolerance for each modification.

To remove all of the modifications from the AA Modif. Details table, click anywhere but on the selected names. To deselect one modification at a time, hold down the Ctrl key and click on the name.

- Name: The abbreviation for the amino acid modification.
- Type: What combinations should be considered for the given modification?

  Fixed: This modification occurs for every instance of the modifiable amino acid in the protein sequence. These modifications are practically stoichiometric, meaning that they are supposed to be present for every modifiable amino acid. All of the targeted amino acids in a peptide are modified and altered by the expected change in mass. It is appropriate to select Fixed Modifications depending on the experimentally applied modifications. For example, you should select Cys_CAM as a Fixed Modification when your sample has been alkylated by iodoacetamide.

  Variable: This modification may or may not have occurred for every instance of the modifiable amino acid in the protein sequence. These

modifications are sub-stoichiometric, meaning that they can be either present or absent for a modifiable amino acid. All permutations are considered. Variable modifications should be used carefully. Since the calculations are more complex, they will increase both the computation time and the false positive rate. Note that many modifications annotated in the UniProt_Swiss-Prot database that alter an amino acid with a discrete mass are considered as variable modifications and are searched by default. They are reported in the results even if a modification was not specified on the Submission page.

- Tolerance: Influence the number of theoretical peptides produced by a modification.

  This is the drop-down list for Fixed Modifications. Depending on your selection, you can reduce the calculation time and the false positive rate.

  All: Each targeted amino acid is modified.

  >=n-1: The number of modifiable amino acids per peptide is all or all except one.

  >=n-2: The number of modifiable amino acids per peptide is all or all except one or two.

  >=n-3: The number of modifiable amino acids per peptide is all or all except one, two or three.

  >=n-4: The number of modifiable amino acids per peptide is all or all except one, two, three or four.

  This is the drop-down list for Variable Modifications.

  None: Each targeted amino acid can potentially be modified.

  <=1: Up to one amino acid modification per peptide is considered.

  <=2: Up to two amino acid modifications per peptide are considered.

  <=3: Up to three amino acid modifications per peptide are considered.

  <=4: Up to four amino acid modifications per peptide are considered.

Enzyme: Choose one of the commonly used reagents to digest proteins. Select Do Not Cleave if no enzyme was used during sample preparation. The term enzyme is defined as any proteolytic cleavage agent (endo or

exoprotease) or chemical agent. The cleavage rules of the listed enzymes are available in the Management Console. Click on the **Def. (Definitions)** link to go to the Management Console page. Click on Cleavage Enzymes to view or edit the definitions. Enzymes can be specific to the user (My defs) or common to all users (Group defs). The table below displays the basic rules for some default enzymes.

Name	**Cleaves where ?**	**Exceptions**
ArgC	C-terminal side of R	if P is C-term to R
ChemDigest_and_Trypsin	C-terminal side of D or K or R and N-terminal side of D	
ChymoTrypsin_(FYL)	C-terminal side of F or Y or L	if P is C-term to F or Y or L
ChymoTrypsin_(FYLW)	C-terminal side of F or Y or L or W	if P is C-term to F or Y or L or W
DoNotCleave	means Does Not Cleave at All	
GluC_(bicarbonate)	C-terminal side of E	if P or E is C-term to E
GluC_(phosphate)	C-terminal side of D or E	if P or E is C-term to D or E
LysC	C-terminal side of K	
Pepsin_(pH>2)	C-terminal side of F or L or W or Y or A or E or Q	
Pepsin_(pH_1.3)	C-terminal side of F or L	
Proteinase_K	C-terminal side of A or C or G or M or F or S or Y or W	
Trypsin_(KR)	C-terminal side of K or R	

Name	Cleaves where ?	Exceptions
Trypsin_(KR_noP)	C-terminal side of K or R	if P is C-term to K or R

Remark: Note that the selection of the DoNotCleave enzyme means that there will be no theoretical digestion of the proteins (except the Methionine processing). Thus this enzyme should be selected if the proteins you are looking for are already present in the database as individual entries.

Missed Cleavage(s): Allowed number of sites (targeted amino acids) per peptide that were not cut. This is an error allowance for enzyme inefficiency (partial cleavage). A complete digest has zero (0) missed cleavages. Any selection other than zero will lengthen the computation time. As the number of potential peptide matches increases, a search is rendered more ambiguous.

- 0: Means that all sites were cleaved as theoretically expected. Lessens the likelihood of random matches.
- 1: All combinations are computed for one uncleaved site (including the case when zero missed cleavages occurred).
- 2: All combinations are computed for two uncleaved sites (including the cases when zero and one missed cleavages occurred).
- 3: All combinations are computed for three uncleaved sites (including the cases when zero, one and two missed cleavages occurred).

Cleavage Mode: Did enzyme digestion occur according to the cleavage rules on one or both ends of the protein? This is an error allowance for enzyme inefficiency.

- Normal: Enzyme digestion occurred specifically on both termini of the protein as defined by the cleavage rules. The cleavage rules defined in the Management Console are strictly applied.
- Half cleaved: Enzyme digestion occurred at just one end specifically and the other end nonspecifically (either the C- or N-terminus). The enzyme cleavage rule defined in the Management Console is applied strictly to one terminus while a nonspecific component is applied to

the other terminus. For every given peptide, Phenyx will generate a set of peptides with identical N-termini but different C-termini and a set of peptides with identical C-termini but different N-termini.

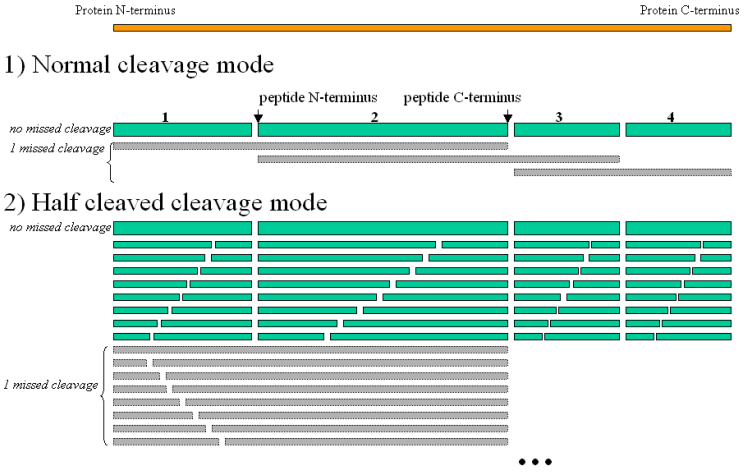


FIGURE 1. Schematic display of the normal cleavage mode (1) versus the half cleaved mode (2). Each block represents a peptide generated *in silico* from a given protein sequence. The larger the number of missed cleavages, the more complex the scheme to express the increased number of generated peptides.

Turbo: A procedure that accelerates a search by pre-processing the data before submitting it to the main scoring calculation. A minimum percentage (20% by default) of the peptide sequence coverage by b^+ (b), b^{2+} (b++), y^+ (y) or y^{2+} (y++) fragment series is looked for. If this percentage is not attained, the spectrum is not submitted for further scoring. If the feature is activated, click on the red arrow to view or edit the default settings.

Conflict Resolution: A conflict can occur during scoring when a mass spectrum matches more than one peptide sequence in the selected protein databases. A given spectrum should ideally correlate to a unique molecular structure, except if a spectrum represents a mixture of peptides. When more than one peptide reasonably matches a MS/MS spectrum should the scoring algorithm deal with them?

- Yes: Phenyx can distinguish all possible matches and decide which is/are the most probable.
- No: Phenyx simply reports all sequences matching with an acceptable score. Then when the results are available, you deal personally with any questionable spectra having multiple matches.

Parent Error Tolerance: There are invariably mistakes or uncontrollable factors that affect measurements. You can allow a certain amount of deviation between the experimental (observed) parent ion masses and the theoretical (calculated) masses. Input the value and choose the corresponding unit in Daltons (Da) or parts per million (ppm).

Acceptance parameters

Click on the red arrow to view or edit all the acceptance parameters.

Minimum Peptide Length: Peptides with less than the specified number of amino acids are reported in the peptide match results but they do not contribute to the protein score. They are labeled as invalid.

Minimum Peptide z-Score: Otherwise known as the peptide or match score. The distribution of calculated scores is compared to that of random peptide sequences in order to find the mean and variance. The z-Score is then a measure of how far and in what direction the score deviates from the distribution's mean.

Maximum Peptide p-Value: The probability of a peptide match in a database occurring by chance with this score or better. A p-Value has a maximum of 1.0. Default score is set to 1e-05. The lower the p-Value, the more significant the match.

AC Score: Minimum significant value for a protein’s accession number score. The AC score is the sum of the best scores for validated peptide sequences. Protein matches scoring lower than this value are rejected from the identified proteins. (Valid peptides from rejected proteins are however also listed in the Compounds Overview page, but AC column is then empty).

Peak lists

File Format: Only one file format can be opted for in a given submission file.

- .btdx: File format defined by Bruker Daltonics.
- .dta: File format defined by SEQUEST/Thermo Electron.
- .mgf: File format defined by Matrix Science.
- mzData: File format defined by HUPO PSI.
- mzXML: File format defined by the Institute for Systems Biology.
- .pkl: File format defined by Waters Corporation.

Peak List(s) : Browse your computer's hard disk to find the file containing the peak lists to be determined. A peak list is a delimited file format that includes the mass-to-charge and intensity for each peak in a mass spectrum. Add one file at a time. Phenyx automatically merges the selected files into a single, continuous file. You can also add several files that were previously compressed in ZIP, GZIP or TAR.GZ format (i.e. one file with a .zip, .gz, .tgz or tar.gz extension).

Resubmitted Peak List(s): If you open the Submission page by using the Resubmit function for a selected job on the Phenyx Desktop, the original data appears here and is automatically uploaded unless the corresponding check box is deselected. You can also add new data with a different format.

Submit: Click on the Submit button to send the submission file to the Phenyx Calculation Unit for processing. A progress bar appears to display the process. The user can effectively resubmit the search by successively clicking the Submit button the number of times desired.

Server log

A message in green characters will appear after the completion of the submission process. For example, Job submission SUCCESSFUL (# 6024) on Mon Apr 24 16:55:16 CEST 2006.

The message disappears and the page becomes available again for a new submission by clicking anywhere in the Submission Page.

Management Console page

Updated: May 10, 2008

The Management Console is an interactive page where users can do the following:

- Create, edit or view enzyme cleavage rules
- Create, edit or view amino acid modification rules
- Create or edit a FASTA sequence database
- Restore an archive of job files
- Import job results from third party software (Mascot, SEQUEST-Bioworks, X!Tandem)
- Access a peaklist format conversion and filtering tool
- Produce various reports
- Generate Phenyx files

The options provided on the Management Console page are dependent upon the roles associated with the account a person is subscribed to. Each account type has a related set of roles and groups. For example, the guest account has anonymous roles and gives non-registered users basic rights. An administrator can perform higher level tasks and generally manages a locally installed Phenyx system for a network of users (i.e. a local group).

User type	**Role possibilities**	**Group possibilities**
Guest	Anonymous	Default
User	User	Default Local Group(s)
Administrator	Admin	Default Local Group(s)

A registered user has write access to the definitions of his/her own cleavage rules and amino acid modifications (My defs). Every user has read access to the definitions of the default group and users with a locally installed version of Phenyx will also have read access to the definitions of their local group(s) (Group defs).

My defs

The My Definitions menu refers to enzymes and amino acid modifications that are specific to the individual logged in. The tools and views are available to all except guest accounts.

Cleavages Enzymes: Use this tool to create, edit, test or delete an enzyme rule. A list of your enzymes appears in the left-hand scrollable box. Click on an enzyme name to select it and view the details.

The following information is applicable to cleavage rules:

- **Name**: A description for the cleavage rule (i.e. an official or common enzyme name) that appears on the Submission page.
- **Cleavage Site**: Describes the type and location of the peptide bonds to be cleaved.

  Cleav At: Lists the targeted amino acids. For example, RK refers to cleavage at arginine and lysine (with no commas between listed amino acids).

  Adjacent: Lists the required amino acids that must be present or absent in order for a bond to be cleaved or not. For example, P means that proline must follow the targeted amino acid in order for a C-terminal cleavage to occur. A caret (^) is used to indicate an absence. It effects all the amino acids that follow (with no commas between listed amino acids). ^PE means proline or glutamic acid must not follow the targeted amino acid in order for a C-terminal cleavage to occur.

  Terminus: Determines if the cleavage occurs on the C- or N-terminal side of the targeted amino acid.

  Regular Expression: Uses a a set of symbols to describe the cleavage rule. For example, (?<=[RK])(?=[^P]) means to cleave on the C-terminal side of R and K except if P occurs C-terminal to arginine or lysine.

- **Gain (Formula)**: Gives the chemical formula to be added to the N-terminus and C-terminus of the peptide fragments. Note that cleavage site is defined as cutting the C-N bond of the peptide bond between two amino acids in a protein sequence.

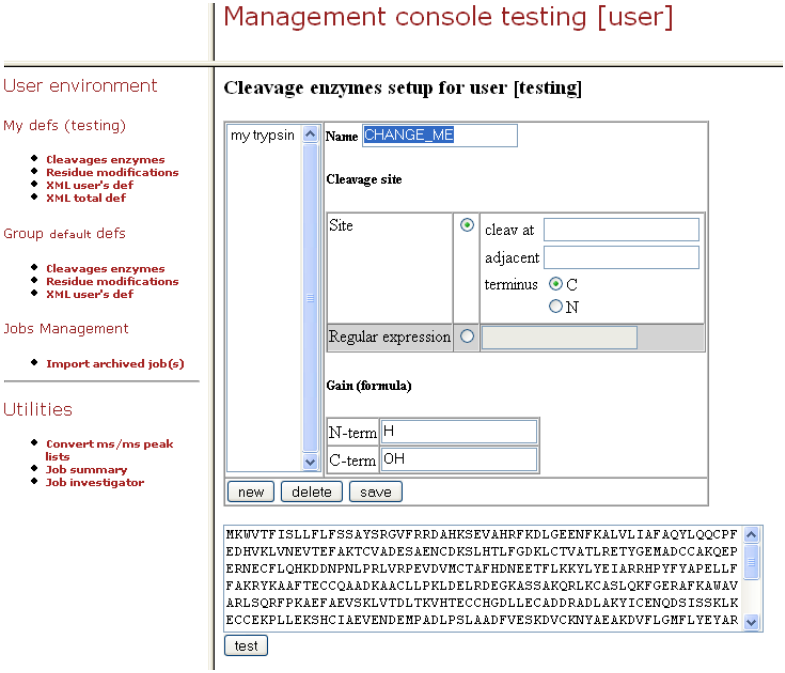
Any registered user can define their own enzyme rules, belonging exclusively to the individual. Several online resources can assist you in creating these rules:

- ENZYME: Available through the Swiss Institute of Bioinformatics (SIB) at http://www.expasy.org/enzyme/.
- Biochemical Pathways: Boehringer Mannheim's wall charts found interactively at http://www.expasy.org/tools/pathways/.
- BRENDA: Available through the Cologne University Bioinformatics Center at http://www.brenda-enzymes.info/.
- KEGG: Kyoto Encyclopedia of Genes and Genomes found at http://www.genome.jp/kegg/.
- Enzyme Nomenclature: Recommendations from the International Union of Biochemistry and Molecular Biology (IUBMB) found at http://www.chem.qmul.ac.uk/iubmb/enzyme/.
- BioCarta: Charting Pathways of Life found at http://www.biocarta.com/.

Define a new enzyme

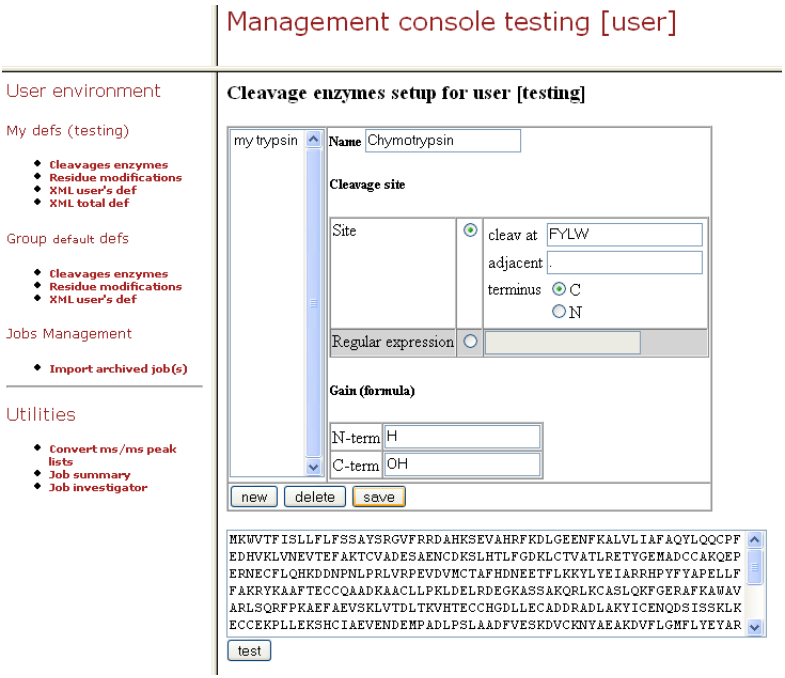
1. Starting from the Phenyx Desktop, click on the Management Console link. The Management Console page opens in a new window.
2. Click on the Cleavages Enzymes link under the My Defs menu. A form appears in the right-hand pane.
3. Click on the New button.

4. Fill in the Name field (e.g. type Chymotrypsin).



5. Fill in the Cleav At field (e.g. type FYLW).
6. Fill in the Adjacent field, if necessary.
7. Select the Terminus (e.g. C).
8. Fill in the Gain (Formula) fields (e.g. type H for hydrogen in the N-term box and type OH for a hydroxyl group in the C-term box).

9. Click on the Save button. The new enzyme is added to the scrollable list of user-defined enzymes on the left-hand side of the pane.



10. Try out the rule on a model protein by clicking on the Test button. The predicted fragments cut by the enzyme are shown at the bottom of the pane.

11. In order for the enzyme to appear on the Submission page, log out and re-log in to the Phenyx Web Interface.

Edit a user-defined enzyme

1. Select an enzyme name in the scrollable list to view its cleavage rule (e.g. Chymotrypsin).

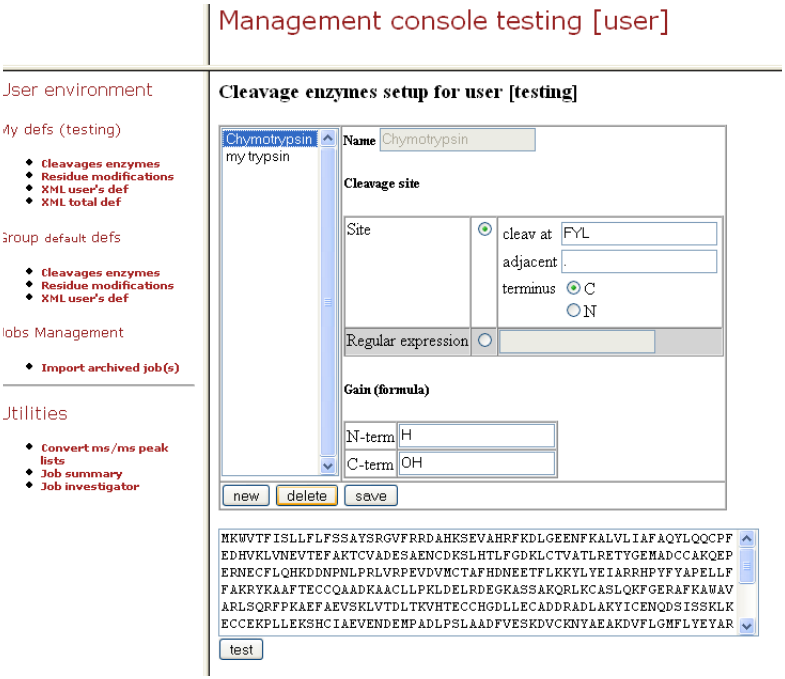
2. Make the desired changes (e.g. modify the Cleav At field from FYLW to FYL).

3. Click on the Save button.

4. Try out the rule on a model protein by clicking on the Test button. The predicted fragments cut by the enzyme are shown at the bottom of the pane.
5. In order to implement the updated enzyme, log out and re-log in to the Phenyx Web Interface.

Delete a user-defined enzyme

1. Select an enzyme name in the scrollable list (e.g. Chymotrypsin).
2. Click on the Delete button. Chymotrypsin is removed from the list of user-defined enzymes.



3. In order for the enzyme to disappear from the Submission page, log out and re-log in to the Phenyx Web Interface.

Residue Modifications: Use this tool to create, edit, test or delete an amino acid modification rule. A list of your modifications appears in the

left-hand scrollable box. Click on a modification name to select it and view the details.

- **Name**: An abbreviation that represents the modification on the Submission page.
- **Description**: An explanatory statement about the modification such as its official or common name.
- **Position**: Describes the type and location of the modified amino acid(s).

  Residue: The standard one-letter code for the targeted amino acid(s).

  Regular Expression: Specifies the rules for a set of strings that you want to match to a pattern.

  Peptide Terminus: Indicates whether the modification occurs on the N-terminal or C-terminal side of the peptide.
- **Mass Modifications**: Describes the mass difference between the unmodified and modified amino acids.

  Monoisotopic Delta Mass: The monoisotopic mass that should be added/subtracted due to the modification. A plus sign (+) or nothing is used to denote an addition. A minus sign (-) is used to denote a subtraction.

  Average Delta Mass: The average mass that should be added/ subtracted due to the modification. A plus sign (+) or nothing is used to denote an addition. A minus sign (-) is used to denote a subtraction.

  Formula: The molecular formula that gives the total number of atoms of each element that are added/subtracted due to the modification. For example, CSH2 adds 1 C, 1 S and 2 H to the targeted amino acid.
- **UniProt Annotations** (i.e. database annotation): Describes known regions or sites of interest in a protein sequence. If annotations are specified in the installed databases, the FTMask corresponds to the term in the header of the database entries.

Any registered user can define their own amino acid modifications, belonging exclusively to the individual. Several online resources can assist you in creating these rules:

- FindMod: Available through the Swiss Institute of Bioinformatics (SIB) at http://www.expasy.org/tools/findmod/.
- RESID: Available through the European Bioinformatics Institute (EBI) at http://www.ebi.ac.uk/RESID/.
- UniMod: Available through the UniMod public database web server at http://www.unimod.org/.
- Delta Mass: Available through the Association of Biomolecular Resource Facilities (ABRF) at http://www.abrf.org/index.cfm/dm.home.

Define a new modification

1. Starting from the Phenyx Desktop, click on the Management Console link. The Management Console page opens in a new window.
2. Click on the Residue Modifications link under the My Defs menu. A form appears in the right-hand pane.
3. Click on the New button.
4. Fill in the Name field (e.g. type DOPA).
5. Fill in the Description field (e.g. type 3,4-Dihydroxy-Phenylalanine).
6. Fill in the Residue field (e.g. type F). If many amino acids are targeted, then simply list them without commas.
7. If needed, then select the N-term or C-term to specify the position of the targeted amino acids on the peptide.
8. Enter the Monoisotopic Delta Mass (e.g. 15.994).
9. Enter the Average Delta Mass (e.g. 16).
10. Click on the Save button. The new modification is added to the scrollable list of user-defined modifications on the left-hand side of the pane.
11. Try out the rule on a model protein by clicking on the Test button. The modified amino acids predicted to occur are highlighted in the sequence at the bottom of the pane.
12. In order for the modification to appear on the Submission page, log out and re-log in to the Phenyx Web Interface.

Edit a user-defined modification

1. Select a modification name in the scrollable list to view its rule (e.g. DOPA).
2. Make the desired changes (e.g. modify the Monoisotopic Delta Mass from 15.994 to 15.9949).
3. Click on the Save button.
4. In order to implement the updated modification, log out and re-log in to the Phenyx Web Interface.

Delete a user-defined modification

1. Select a modification name in the scrollable list (e.g. DOPA).
2. Click on the Delete button. DOPA is removed from the list of user-defined modifications.
3. In order for the modification to disappear from the Submission page, log out and re-log in to the Phenyx Web Interface.

XML User's Def: The Extensible Markup Language file showing the statements used by Phenyx to define the user's enzyme and modification rules.

XML Total Def: The Extensible Markup Language file showing all the statements used by Phenyx to define both the current user's and the affiliated group's enzyme and modification rules. It also includes XML tags describing the elements, molecules, amino acids, nucleic acids, fragment ions, etc. A non-exhaustive list of Phenyx XML tags is given below. Most nested tags are not described.

XML Tag	**Description**
<inSilicoDefinitions>	Definitions used in the Phenyx calculations
<elements>	Defines the chemical elements
<isotopes>	Defines a chemical element with the same atomic number but different atomic mass
<molecules>	Defines common molecules such as water

XML Tag	Description
<codons>	Defines the standard nucleic acids
<aminoAcids>	Defines the standard amino acids
<mRNAcodons>	Defines a specific sequence of three consecutive nucleotides which are part of the genetic code. Messenger RNA are produced by transcription and used during translation as a template for protein synthesis
<cleavEnzymes>	Defines the enzymes that cleave proteins
<fragTypeDescriptions>	Defines the dissociation of a protein into peptide fragments in a mass spectrometer
<series>	Defines the standard fragment ion series (a, b, c, x, y and z)
<losses>	Defines the standard loss of water or ammonia
<fragTypes>	Defines the different possible combinations of fragments (where ion series, charge state and loss of water and ammonia are the variables)
<internalFragTypes>	Defines possible fragments formed when peptide bonds are cleaved on both the N- and C-termini
<modRes>	Defines the modifications that change standard amino acids
<site>	Defines the position of the amino acids to be affected by the enzyme or modification rule
<sprotFT>	Defines what annotations to look for in the Swiss-Prot feature table

Group defs

The first Group Definitions menu will refer to the default enzymes and amino acid modifications that are automatically displayed to all users. You may see additional Group Definitions menus that refer to the enzymes and amino acid modifications that are only displayed to users in a local group. These views are available to all registered users. Please contact GeneBio at support@phenyx-ms.com with questions or comments about any predefined settings.

Account Settings

Change Password: This tool allows you to change your Phenyx password. Enter the new password in the first box. Confirm the new password by retyping it in the second box. Click the Change button.

Databanks

Private Databank: This tool enables the user to add and refine new sequence entries that are FASTA formatted. Copy and paste the sequence(s) in FASTA format into the text box. Click on the Load button. In order for the user's databank (called privatedb_username) to appear in the list of Database(s) on the Submission page, log out and re-log in to the Phenyx Web Interface. Only licensed users of Phenyx are able to create their own repository of FASTA sequences. Please contact GeneBio at support@phenyx-ms.com to temporarily activate this feature for a trial period.

Jobs management

Browse files or export job results

browse files

This tool generates a complete look at the Phenyx files associated with a job. Select a job number. Click on the Browse Files button. The directory of all Phenyx files for the job appears at the bottom of the page. By clicking on a file name, you can either open or save the file to your hard disk. With Mozilla Firefox, you can save the file by right clicking and

choosing This Frame > Save Frame As in the pop-up menu. With Internet Explorer, you can copy and paste the file into a text editor by right clicking and choosing Select All in the pop-up menu.

export

This tool exports the specified Phenyx file for selected jobs (use the Shift or Ctrl keys to make multiple job IDs selection when possible regarding the selected export feature). Select the type of file from the drop-down list. Click on the Export button.

See also:

http://gbwiki.genebio.com/mediawiki/index.php/CGI_export

Below are some details about two of the export features:

false discovery rate pept matches (excel)**:** In order to get information about the FDR (false discovery rate) in your results, you have to perform two jobs with the same data, one job searching on the "true database" (dbtrue argument), the other one searching on the "false database" (dbfalse argument). You can also perform only one search with both databases selected in the search field. The false database is a random database generated from the original database. On the public server the random database corresponding to uniprot_sprot is named uniprot_sprot_rev database (reverse sequences). GeneBio prepares forward and random databases for the licensed users, please contact GeneBio at support@phenyx-ms.com to have personal database seting for a trial period.

1. select "false discovery rate pept matches (excel)" in the drop-down menu
2. select job IDs (use the Shift or Ctrl keys to make multiple selections)
3. paste the extra arguments (for instance --dbtrue=uniprot_sprot --dbfalse=uniprot_sprot_rev)
4. click on the export button to get the excel FDR report

unmatched fragment spectra list(mgf) format: This export generates a file (mgf format) that only contains compounds for which there is no peptide interpretation.

Import Phenyx jobs or jobs from other search engines

Import

This tool converts data such as Mascot, SEQUEST-Bioworks, X!Tandem or X!Hunter output as well as Phenyx archives into jobs with new identification (ID) numbers.

See also

http://gbwiki.genebio.com/mediawiki/index.php/CGI_import

- Select the relevant protein identification software in the Import from drop-down menu.
- If working with saved Results data, then click on the Browse button to locate the Local File or archive on your hard disk.

Otherwise, copy and paste the Remote URL of in the Add URL box.
If the data are only accessible using a login and password, then type: http://login:password@result_url instead of a basic http://result_url.

- Click on the Import button to convert the file into a job with a new identification (ID) number.
- Go to the Phenyx Desktop to start working with the job.

In the Title column of the Job Report, an information about the origin of the imported job is given, for example the original ID of an imported Phenyx job is referenced [in brackets]. In case of an imported Mascot job, the Title column displays [imported from mascot] and file title. In the Comment column of the Job Report, information about the imported file or URL is given.

Utilities

Convert MS/MS Peak Lists: This tool allows to transform one type of MS/MS peak list into a different file format. You can also filter the original peak list. Click on the question mark (?) links on the on-screen form for further explanations. Please contact GeneBio at support@phenyx-ms.com with questions or comments about any specific needs.

Status page

Updated: May 10, 2008

The Status page or user's Desktop allows you to monitor the state of running, pending or failed jobs.

If a job is currently running, then a progress bar at the bottom of the page indicates the two main cpu time-consuming activities, i.e. the DatabaseN_spectra_rnd and the DatabaseN_DB_blocks processes.

If a job completes normally, then you are automatically redirected to the Proteins Overview page.

The comments explaining the status of a given job number are described below:

- Reading source data: The job is submitted.
- Writing idj: The submission parameters and peak lists are transformed into a Phenyx XML file.
- Compressing input file: Data is encoded to take up less storage space and less bandwidth for transmission.
- Submit to Database N: The specified database is queried.
- DatabaseN_spectra_rnd: Status of the random search of the MS/MS spectra expressed as a percentage.
- DatabaseN_DB_blocks: Status of the peptide matching procedure expressed as a percentage.
- Apply post processing command: The results are transformed into a Phenyx XML file.
- Completed: The job finished successfully.
- Error: The job failed to process.
- Pending: The job is in the queue awaiting processing.

If an error occurs, you can find out why a job failed and perhaps attempt to fix the problem yourself by clicking on the Comment link to go to the Troubleshooting page.

GeneBio's technical support is of course at your disposal should you have question or comment (support@phenyx-ms.com).

Once a job reaches the point at which the input file is compressed, you are able to Resubmit or view the Parameters for that job. Simply click on its ID in the Job Report on the Phenyx Desktop in order to open the Job Menu and perform these operations.

Parameters page

Updated: May 10, 2008

The Parameters page contains a concise restatement of the main submission specifications including the unique job identification number, specified title, peak list file names, databases, accession numbers, scoring model, modifications, thresholds, etc.

You can also use the Parameters option on the Phenyx Desktop in order to see a comparison of submission settings for different jobs. Select the desired job(s) by checking their boxes in the Job Report. Choose Parameters from the Actions on Selected drop-down menu. The selected job(s) are inserted in a table on the Parameters page.

Note that currently Parameters are not accessible for imported job from other search engines.

Proteins Overview page

Updated: May 10, 2008

This first page of the results gives an overview of the identified proteins.

Parameters: Click on the Parameters link to go to a concise restatement of the main submission points including the specified title, scoring model, modifications, thresholds, etc.

Compounds Overview: Click on the Compounds Overview link to go to a listing of the mass spectra and corresponding peptide matches.

The Proteins Overview is divided into three panes: Best-scoring Protein Table (top left), Protein Details (top right) and Peptide Match Table (bottom). With Mozilla Firefox, two of the panes are resizable by clicking and dragging the Expand/Contract boxes. Horizontal and/or vertical scroll bars appear if needed to view the information.

Best-scoring protein table

The proteins are listed according to their accession numbers and sorted in descending order according to score (the protein with the highest score is the best match). When you click on a row in the table, it is highlighted. By selecting a row, more details about the protein are presented to the right of the table. A summary of the matched peptides for the selected protein is given below the protein table.

AC: Database accession number. The suffix _WOSIG0 denotes the protein sequence without the Signal sequence. The suffix _WOPP0 denotes the Peptide 1 in SwissProt. The suffixes _CHAIN0 or _VAR1 denote Chain 1 from the original protein entry or Variant 2 from the original protein entry, respectively.

ID: Database protein identification code. Swiss-Prot refers to this as the entry name whereas other databases use the accession number.

Score: Effective score, the protein score that is recalculated based on user-validated peptides. It is the sum of the best scores for validated peptide sequences.

 # Peptides: Peptide numbers, the number of valid peptide matches followed by the total number of peptide matches found for the given protein.

Cov.: Coverage, the percent ratio of all amino acids from valid peptide matches to the total number of amino acids in the protein.

Description: Official protein name followed by common synonyms (in parentheses).

If a protein shares all of its validated peptides with another protein, then it is considered to be a subset and will not appear in the best-scoring list. It appears in the Protein Details pane under Subset for the principal and better-scoring protein.

Protein details

The main points about the protein are supplied on the right-hand side of the best-scoring protein table. The brief statement includes the identification code (ID), the accession number (AC), the database name and the official protein name followed by common synonyms (in parentheses). Click on the AC link to go to the actual database entry. For example, a protein identified in UniProt will open a page from the ExPASy Proteomics Server developed by the Swiss Institute of Bioinformatics (SIB). This link works for public databases such as UniProt Swiss-Prot and TrEMBL, but not for databases which are not widely accessible such as on the Web. Please contact GeneBio at support@phenyx-ms.com for further assistance. Click on the Protein Details link to go to the Protein Match Details page.

Subset: A list of all proteins with peptide matches that overlap with those of another identified protein. The listed proteins have peptides that are included in the set of matches for a protein with a better score. Only the better-scoring proteins are considered as significant results and appear in the best-scoring protein table. Click on the AC link to go to the

actual database entry. For example, a protein identified in UniProt will open a page from the ExPASy Proteomics Server developed by the Swiss Institute of Bioinformatics (SIB). Click on the Protein Details link to go to the Protein Match Details page.

Peptide match table

A peptide match is the pairing of an experimental fragmentation spectrum to a theoretical segment of a protein. A summary of all matched peptides for the selected protein is given.

Auto: A plus sign (+) means that the peptide match is considered valid by Phenyx and computed in the final scoring of the protein. A minus sign (-) means that the peptide match did not fulfill the specified p-Value thresholds or has been rejected after conflict resolution (if activated) and therefore is not retained in the final scoring.

User: The validity of the peptide match as determined by the user on the Compounds Overview page. A plus sign (+) means that the peptide match is considered valid by the user and computed in the final scoring of the protein. A minus sign (-) means that the peptide match is considered invalid by the user and therefore is not retained in the final scoring. If the user chooses not to manually validate peptide matches, then the validity assigned by Phenyx appears by default in the User column.

Sequence: Amino acid sequence of the matched peptide. The one-letter amino acid codes are used. Peptide sequences that include the original N- or C-terminus of the protein are denoted by a minus one (-1) and forward slash (/). For example, K/IIEEDDAYDFSTDYV/-1 indicates that V is the C-terminus of the protein from which the peptide originated. Internal peptide sequences are identified by the use of forward slashes (/) and the original protein amino acids that come before and after the peptide are given. K/KLVLILNK/S is an internal peptide composed of the amino acids KLVLILNK (K and S are the amino acids that appear before and after the peptide in the protein). Bolded amino acids were modified.

Click on the Sequence link to go to the Peptide Match Details page.

Search: By clicking on the BLAST link, the SIB (Swiss Institute of Bioinformatics) BLAST Network Service opens in a new browser window. This tool enables you to find similar entries to the selected sequence in ExPASy's protein and nucleotide databases.

z: Charge state of the theoretical peptide.

m/z: Experimental mass-to-charge ratio.

d m/z: Delta mass-to-charge, the value is the difference between the theoretical m/z of a matched peptide and the observed m/z of the parent ion.

z-Score: The distribution of calculated scores is compared to that of random peptide sequences in order to find the mean and variance. The z-Score is then a measure of how far and in what direction the score deviates from the distribution's mean.

p-Value: The probability of a peptide match in a database occurring by chance with this score or better. The lower the p-Value, the more significant the match.

Position: The numerical start and end of the peptide in the main protein sequence.

 # MC: The number of missed cleavages, the allowed number of sites (targeted amino acids) per peptide that were not cut. This is an error allowance for enzyme inefficiency (partial cleavage).

- 0: Means that all sites were cleaved as theoretically expected. Lessens the likelihood of random matches.
- 1: All combinations are computed for one uncleaved site (including the case when zero missed cleavages occurred).
- 2: All combinations are computed for two uncleaved sites (including the cases when zero and one missed cleavages occurred).
- 3: All combinations are computed for three uncleaved sites (including the cases when zero, one and two missed cleavages occurred).

- 4: All combinations are computed for four uncleaved sites (including the cases when zero, one, two and three missed cleavages occurred).
- Half: Enzyme digestion occurred at just one end of the protein specifically and the other end nonspecifically (either the C-terminus or the N-terminus).

Modif.: Modification(s), any amino acid undergone are cataloged. The colon (:) is a counter that surrounds each amino acid, thus differentiaing the N- and C-terminuses from the side chains of the starting and ending amino acids. For example, a sequence of 8 amino acids will be denoted by 9 colons. For a peptide MSGECACK, the code :Oxidation_M:::::::: represents the oxidation of the side chain of methionine. Find additional examples below. The detailed description of what modification occurs is also shown using this particular nomenclature on the Protein Match Details page.

Scenario	Nomenclature	Sequence example
Unmodified sequence	::::::::	AQQAADK
Modified N-terminus of peptide	ACET_nterm::::::::	(ACET_nterm)AQQAADK
Modified side chain of N-terminal amino acid	:Oxidation_M::::::::	M(Oxidation_M)SGECACK
Modified amino acid in an internal peptide sequence	::::Cys_CAM:::::	FSSC(Cys_CAM)GGSK
Modified side chain of C-terminal amino acid	:::::::Oxidation_HW:	ATAGDTH(Oxidation_HW)

Scenario	Nomenclature	Sequence example
Modified C-terminus of peptide	:::::::::METH_cterm	DFLMLYAR(METH_cterm)

Compound: This is a label commonly attached to LC MS/MS peak lists. The description is used for identification purposes. It can contain a Cmpd (compound or LC peak) number, +MSn (abbreviation to indicate that the spectrum was obtained in MS/MS mode), the mass-to-charge ratio assigned to the LC peak in parentheses and the LC retention time in minutes. The various peak list formats generate different compound descriptions. For example, the spectrum title is inserted into this field for a .mgf file. For both .dta and .pkl files, numbers are assigned to consecutive MS/MS spectra and these values appear in the Compound column. In order to read the full description in the Status Bar at the bottom of the Web page, move your mouse cursor over the truncated compound. If this feature is not activated in your Internet browser, then select Status Bar in the View menu.

Click on the Compound link to go to the Compounds Overview page.

Protein Match Details page

Updated: May 10, 2008

ID: Database protein identification code. Swiss-Prot refers to this as the entry name whereas other databases use the accession number.

AC: Database accession number (in parentheses). The suffix _WOSIG0 denotes the protein sequence without the Signal sequence. The suffix _WOPP0 denotes the Peptide 1 in SwissProt. The suffixes _CHAIN0 or _VAR1 denote Chain 1 from the original protein entry or Variant 2 from the original protein entry, respectively. Click on the AC link to go to the actual database entry. For example, a protein identified in UniProt will open a page from the ExPASy Proteomics Server developed by the Swiss Institute of Bioinformatics (SIB).

DB: Database description.

Description: The description line from the sequence database that includes the official protein name, for example.

Synonyms: Other common names for the protein (in parentheses).

Protein information table

This table displays information on the current protein and is updated according to the manual validation.

Score: Effective score, the protein score that is recalculated based on user-validated peptides. It is the sum of the best scores per valid peptide sequences.

#Peptides: Peptide numbers, the number of valid peptide matches followed by the total number of peptide matches found for the given protein.

Cov: Coverage, the percent ratio of all amino acids from valid peptide matches to the total number of amino acids in the protein.

Peptide match table

A peptide match is the pairing of an experimental fragmentation spectrum to a theoretical segment of a protein. Only peptides that respect the z-Score threshold are reported. A summary of all matched peptides for the selected protein is given.

Auto: A plus sign (+) means that the peptide match is considered valid by Phenyx and computed in the final scoring of the protein. A minus sign (-) means that the peptide match did not fulfill the p-Value and/or the minimal peptide length thresholds, or has been rejected after conflict resolution (if activated) and therefore is not retained in the final scoring.

User: The validity of the peptide match as determined by the user on the Compounds Overview page. A plus sign (+) means that the peptide match is considered valid by the user and computed in the final scoring of the protein. A minus sign (-) means that the peptide match is considered invalid by the user and therefore is not retained in the final scoring. If the user chooses not to manually validate peptide matches, then the validity assigned by Phenyx appears by default in the User column.

Sequence: Amino acid sequence of the matched peptide. The one-letter amino acid codes are used. Bolded amino acids were modified. Click on the Sequence link to go to the Peptide Match Details page.

Search: By clicking on the BLAST link, the SIB (Swiss Institute of Bioinformatics) BLAST Network Service opens in a new browser window. This tool enables you to find similar entries to the selected sequence in ExPASy's protein and nucleotide databases.

z: Charge state of the theoretical peptide match.

Delta m/z: The value is the difference between the theoretical m/z of a matched peptide and the observed m/z of the parent ion.

Score: Otherwise known as the peptide z-Score. The distribution of calculated scores is compared to that of random peptide sequences in order to find the mean and variance. The z-Score is then a measure of

how far and in what direction the score deviates from the distribution's mean.

p-Value: The probability of a peptide match in a database occurring by chance with this score or better. The lower the p-Value, the more significant the match.

Position: The numerical start and end of the peptide in the main protein sequence.

 # MC: The number of missed cleavages, the allowed number of sites (targeted amino acids) per peptide that were not cut. This is an error allowance for enzyme inefficiency (partial cleavage).

- 0: Means that all sites were cleaved as theoretically expected. Lessens the likelihood of random matches.
- 1: All combinations are computed for one uncleaved site (including the case when zero missed cleavages occurred).
- 2: All combinations are computed for two uncleaved sites (including the cases when zero and one missed cleavages occurred).
- 3: All combinations are computed for three uncleaved sites (including the cases when zero, one and two missed cleavages occurred).
- 4: All combinations are computed for four uncleaved sites (including the cases when zero, one, two and three missed cleavages occurred).
- Half: Enzyme digestion occurred at just one end specifically and the other end nonspecifically (either the C-terminus or the N-terminus).

Modif.: Modification(s), any amino acid modifications undergone are cataloged. The colon (:) is a counter that surrounds each amino acid, thus differentiaing the N- and C-terminuses from the side chains of the starting and ending amino acids. For example, a sequence of 6 amino acids will be denoted by 7 colons. For a peptide MSGECACK, the code :Oxidation_M::::::: represents the oxidation of the side chain of methionine.

Sequence views

Two different graphical representations of the protein sequence and its coverage by identified peptide matches are given at the bottom of the page (Figure 1).

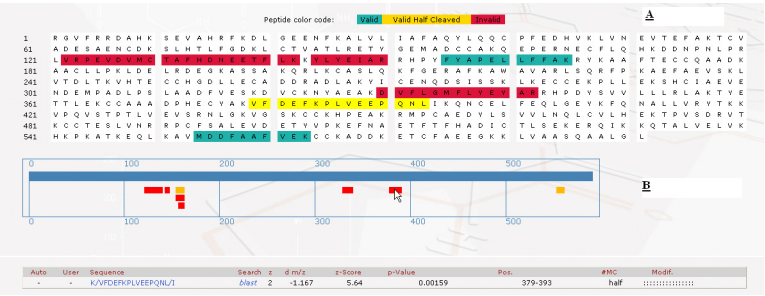


FIGURE 1. The first view (A) is a graphical representation of the protein sequence and the location of the identified peptides. Information about the sequence coverage by the peptide matches can be visualized in the second view (B).

The first view is the ordered arrangement of the amino acids in the protein. The sequence is written from the N-terminus, at the left, to the C-terminus, at the right, using the one-letter amino acid codes. The start position of the first amino acid in each row is given on the left-hand side. The matched peptide sections are highlighted in colored boxes (green for valid peptides and red for invalid peptides).

The second view is an illustration of the number of peptide matches that covered portions of the protein sequence. The start, middle and end positions are labeled across the top of the diagram. The solid bar indicates the total sequence. The peptide matches are listed and highlighted in colored boxes (green for valid peptides, red for invalid peptides and orange for valid half cleaved peptides).

By moving the mouse cursor over one of the peptides in the second view, the sequence is highlighted in yellow in the first view and details about the peptide match are shown at the bottom of the page.

Note that sequences up to 2000 characters are entirely displayed in the Protein Match Details view. When a longer sequence is matched, Phenyx displays and stores the portion of the sequence covered by the valid

peptides. For example, if a protein is 5000 amino acids long and the first valid peptide match is at position 1000 and the last one at position 4000, the view will display positions 1000 to 4000 plus some additional amino acids. If a peptide is matched with an invalid status at position 500, the displayed sequence will still be the same.

Compounds Overview page

Updated: May 10, 2008

Identified peptide matches are tabulated according to their corresponding compounds (i.e. spectra). The matches are visually labeled in several ways in the Compounds Overview. Multiple matches for a given spectrum appear with the letter M in the Multi column. The description of the spectrum is repeated in the Compound column for each of the conflicting peptides. Only peptides that respect the z-Score threshold are reported. A plus (+) or minus (-) sign is displayed in the Auto column depending on match's validity. All peptide matches are also color-coded in the Sequence column according to their definition:

Term	**Validity**	**Color Code**	**Definition**
Valid	plus (+)	White	The peptide match respects the p-Value and minimal peptide length thresholds specified by the user on the Submission page.
Rejected	minus (-)	Gray	The peptide match does not respect the p-Value and/or minimal peptide length thresholds specified by the user on the Submission page.
Conflict winner	plus (+)	Green	A winner is a valid sequence that successfully satisfies the conflict resolution algorithm. Multiple winners are possible for the same spectrum.

Term	Validity	Color Code	Definition
Conflict loser	minus (-)	Red	A loser is a valid sequence that does not successfully satisfy the conflict resolution algorithm. In comparison to the other conflict(s), its value is worse. Multiple losers are possible for the same spectrum.

Multi: The letter M denotes that a conflict exists between the given peptide and other potential matches that were found for the same compound.

Compound: This is a label commonly attached to LC MS/MS peak lists. The description is used for identification purposes. It can contain a Cmpd (compound or LC peak) number, +MSn (abbreviation to indicate that the spectrum was obtained in MS/MS mode), the mass-to-charge ratio assigned to the LC peak in parentheses and the LC retention time in minutes. The various peak list formats generate different compound descriptions. For example, the spectrum title is inserted into this field for a .mgf file. For both .dta and .pkl files, numbers are assigned to consecutive MS/MS spectra and these values appear in the Compound column. In order to read the full description in the Status Bar at the bottom of the Web page, move your mouse cursor over the truncated compound. If this feature is not activated in your Internet browser, then select Status Bar in the View menu.

Auto: A plus sign (+) means that the peptide match is considered valid by Phenyx and computed in the final scoring of the protein. Note that valid peptides can also be *not* associated to a protein if the protein AC Score theshold is not reached. A minus sign (-) means that the peptide match did not fulfill the p-Value and/or the minimal peptide length thresholds, or has been rejected after conflict resolution (if activated) and therefore is not retained in the final scoring.

User: The validity of the peptide match as determined by the user. The assignments made automatically by Phenyx appear by default. Click on

the box to select or deselect the validity. A checked box means that the peptide match is considered valid by the user and should be computed in the final scoring of the protein. An empty box means that the peptide match is considered invalid by the user and therefore should not be retained in the final scoring. Click on the Save button to store your validations.

Note that when a search is done in several databases, Phenyx displays the automatic validity status found in the first database as the User validity status for a given peptide match on the Compounds Overview page. If the match is Auto valid in the first database but not Auto valid in the next database (due to the influence of the different database sizes on the results), the Auto invalid status is reported on the corresponding Proteins Overview and Protein Match Details pages with the User status set as valid.

Sequence: Amino acid sequence of the matched peptide. Bolded amino acids were modified. Click on the Sequence link to go to the Peptide Match Details page.

Search: By clicking on the BLAST link, the SIB (Swiss Institute of Bioinformatics) BLAST Network Service opens in a new browser window. This tool enables you to find similar entries to the selected sequence in ExPASy's protein and nucleotide databases.

z: Charge state of the theoretical peptide match.

d m/z: Delta mass-to-charge, the value is the difference between the theoretical m/z of a matched peptide and the observed m/z of the parent ion.

z-Score: The distribution of calculated scores is compared to that of random peptide sequences in order to find the mean and variance. The z-Score is then a measure of how far and in what direction the score deviates from the distribution's mean.

AC: List of all proteins that correspond to the compound. Click on an AC number to go to the Protein Match Details page. Valid peptides can also be not associated to any protein if the protein is not validated (that

means, the minimal AC score value for the protein, as defined in the Submission page, is not reached)

Functions

Hide User Invalidated Entries: Select the Hide User Invalidated Entries box to remove any unchecked peptide matches in the User column from the table. Deselect the box to redisplay all peptide matches.

Display Following AC: Specify a list of accession numbers to display in the table. More than one AC number can be entered separated by a space. Incomplete AC numbers are accepted such as P00 (no wildcard is necessary). Select the Display Following AC box to update the table. Deselect the box to redisplay all peptide matches.

The Hide User Invalidated Entries and Display Following AC filters are used in combination when both boxes are checked.

Select All: Click on the Select All button to check all the boxes in the User column.

Deselect All: Click on the Deselect All button to uncheck all the boxes in the User column (i.e. none of the peptide matches are selected).

Reset to Auto Validation: Click on the Reset to Auto Validation button to restore the default selections made by Phenyx. Any user validations in memory are lost.

Save: After manually validating peptide matches, click on the Save button to store the current selections. The date and time of this action is logged at the top of the page.

Export Last Saved Validated Peptides: After saving a selection of validations, click on the Export Last Saved Validated Peptides link to export the Phenyx result file in Extensible Markup Language format to a new browser window. Use the Save As option to store a copy of the data to your hard disk.

Validate peptide matches

User validations serve three main purposes: to substantiate peptide identifications by filtering false positives from true positives, to resolve conflicts and to generate scoring models fitted to your experiments.

1. Evaluate a peptide's p-Value and z-Score on the Proteins Overview page (i.e. a reliable match has a low p-Value and high z-Score).
2. When a case is ambiguous, look at the Peptide Match Details page to assess the quality of the mass spectrum and fragment ion table.
3. On the Compounds Overview page, check/uncheck the box in the User column depending upon if the peptide match should be considered as valid or disregarded as invalid.
4. When you are finished validating and invalidating peptide matches, click on the Save button to store the current selections.

The effects of user validations are immediate on the Compounds Overview, Proteins Overview, Protein Match Details and Phenyx Desktop pages. If these pages were already open, then you will need to click on the Internet browser's Refresh button to notice the changes.

- Compounds Overview: The color coding in the Sequence column will change according to your selections.
- Proteins Overview: The values for the Score, # Peptides (number of valid peptides) and Coverage in the Best-Scoring Protein Table will be modified. In other words, selecting or deselecting a peptide match modifies the related protein(s). The new validity sign (plus or minus) will appear in the User column of the Peptide Match Table.
- Protein Match Details: The new validity sign (plus or minus) will appear in the User column of the Peptide Match Table. The color coding in the first sequence view will change according to your selections.
- Phenyx Desktop: The Comment column in the Job Report will update with the number of valid peptide matches (in parentheses).

Peptide Match Details page

Updated: May 10, 2008

A peptide match is the pairing of an experimental fragmentation spectrum to a theoretical segment of a protein.

Compound Description: This is a label commonly attached to LC MS/MS peak lists. The description is used for identification purposes. It can contain a Cmpd (compound or LC peak) number, +MSn (abbreviation to indicate that the spectrum was obtained in MS/MS mode), the mass-to-charge ratio assigned to the LC peak in parentheses and the LC retention time in minutes. The various peak list formats generate different compound descriptions. For example, the spectrum title is inserted into this field for a .mgf file. For both .dta and .pkl files, numbers are assigned to consecutive MS/MS spectra and these values appear in the Compound description.

Sequence: The amino acid sequence of the matched peptide is given in the parentheses. The one-letter amino acid codes are used.

Modifications: Any amino acid modifications undergone are also given in the parentheses. The colon (:) is a counter that surrounds each amino acid, thus differentiaing the N- and C-terminuses from the side chains of the starting and ending amino acids. For example, a sequence of 6 amino acids will be denoted by 7 colons. For a peptide MSGECACK, the code :Oxidation_M:::::::: represents the oxidation of the side chain of methionine.

Theoretical Mass: Calculated mass of the peptide.

Experimental m/z: Experimental mass-to-charge, the observed mass-to-charge ratio of the parent ion.

Experimental Charge: z, observed charge state of the parent ion.

Experimental # Peaks: Number of peaks, the peak count in the mass spectrum.

Match Delta m/z: The value is the difference between the theoretical m/z of a matched peptide and the observed m/z of the parent ion.

Match Score: Otherwise known as the peptide z-Score. The distribution of calculated scores is compared to that of random peptide sequences in order to find the mean and variance. The z-Score is then a measure of how far and in what direction the score deviates from the distribution's mean.

Match Charge: Theoretical z, calculated charge state of the peptide.

Mass spectrum

The plot of peak intensity versus m/z is given.

Zoom Factor: Input a number in order to multiply the magnification of x-axis in the spectrum. Then click on the area of the spectrum that you would like to become the center of the enlarged view.

Ion Series Selection: Select one or more fragment ion types in the table and click on the Apply Changes button. The related peaks are highlighted in the mass spectrum. The peaks are annotated with the corresponding details. The color legend appears under the graph. For selected fragment series, the colors are as follows:

- Red: Unmatched mass spectral peak.
- Blue: Fragment peaks with matching peptides. The blue peaks are plotted using the theoretical m/z and the intensity of the corresponding experimental m/z.

Input a number in the Max Intensity box and click on the Apply Changes button in order to rescale the y-axis.

Reset X & Y axes: Click on the Reset button in order to zoom out and view the full spectrum.

Fragment ion tables

Different tables consolidate the fragmentation findings.

1.-Main table

There is one column per peptide amino acid and one row per fragment series. Bolded amino acids were modified. N-terminal fragments (a, b, c, etc.) and C-terminal fragments (x, y, z, etc.) are the most common products from the dissociation and detection of protonated peptides in a mass spectrometer. The two rows of numbers in the table indicate the position of bond cleavage in the ordered peptide sequence for N-terminal fragments (1, 2, ..., n) and C-terminal fragments (n, ..., 2, 1). This standard nomenclature is based on the original proposal of Roepstorff P and Fohlman J (Biomedical Mass Spectrometry. 1984; 11(11): 601).

Value Types: Use the drop-down list to change the information displayed in the table. The possible views are:

- Delta Mass: The m/z deviations between the experimental and theoretical masses are provided for the related fragments.
- Theoretical Mass: Theoretical m/z are provided for the related fragments.
- Peak Mass: Experimental m/z are provided for the related fragments.
- Peak Intensity: Spectral ion intensities are provided for the related fragments.

The first column of the table defines the fragment.

- Type of fragment: The list changes as a function of the ion series considered during the scoring process and as a function of the match validity.
- Charge state: No plus sign = 1 charge, ++ = 2 charges or +++ = 3 charges.
- Functional group lost: -NH3 = loss of ammonia or -H2O = loss of water. The asterisk (*) indicates that a variable number of molecules is lost dependent upon the amino acids in the peptide. For example, y++-H2O* refers to the doubly-charged y fragments with potential losses of water (1, 2 or 3 molecules can be subtracted because only the amino acids S and T can lose water).

In order to interpret the rest of the table, several representations are used:

- Empty box: No experimental m/z matches the corresponding theoretical fragment.
- Gray box: Impossible fragment. For example, it is not possible to have a N-terminal fragment VFSN with a loss of NH3 as none of these amino acids have an ammonia group.
- Colored box: Match made between the theoretical and experimental m/z. The different colors signify the relative intensity of the fragment (calculated as a percentage of the most intense peaks in the mass spectrum).

Color codes show the percentage of most intense peaks in the mass spectrum that derived from that type of fragment

2.-Neutral losses, immonium ions and precursor masses tables

Below the main table, additionnal tables may display ion matches specific to

- fragments with potential neutral losses
- parent ions with potential neutral losses
- immonium ions.

Description of the ion fragment with potential neutral losses is given, with the corresponding delta mass. Number in brakets refers to the position of bond cleavage in the ordered peptide sequence. The color code for intensity rank is applied.

Description of the precursor is given with potential neutral losses regarding its charges. Color code for intensity rank is applied to report the delta mass.

Immonium ion is labelled with the one-letter code for the corresponding amino acid. Color code for intensity rank is applied to report the delta mass.

Results Comparison page

Updated: May 10, 2008

This page allows the user to view one or more job results in tabular format. The table is organized according to identified proteins. It is easy to do a side-by-side comparison when multiple jobs are selected. Output from other protein identification software such as SEQUEST-Bioworks, Mascot or X!Tandem can also be opened in the table.

Insert jobs in comparison table

There are two ways to create a comparison. The most straightforward possibility is to use the Compare option on the Phenyx Desktop. Select the desired job(s) by checking their boxes in the Job Report. Choose Compare from the Actions on Selected drop-down menu. The selected job(s) are inserted in a table on the Results Comparison page. In addition, their ID numbers and title appear in the Jobs List.

Add Following Jobs: Alternatively, the Add Following Jobs feature on the left side of the Results Comparison page allows you to construct a new table or add jobs to an existing table.

Define a new comparison

1. Starting from the Phenyx Desktop, click on the Result Comparison link. The Results Comparison page opens in a new window.
2. If necessary, reset the page by clicking on the Empty list button.
3. Enter the known job identification number(s) in the Add Following Jobs box. More than one job number can be entered by using a comma (,) and no space to separate them.
4. Click on the Update button to open the results in the table. The ID numbers appear checked in the Jobs List.

In order to insert additional jobs into a table, repeat Steps 2 and 3 above.

Jobs List: A list of ID numbers and titles for the most recently compared jobs is given on the left side of the Results Comparison page. If its box is checked, then the job is currently shown in the comparison table. A job that was previously deactivated can be reinserted into the table by selecting it (box checked) and clicking the Update button. A green V next to an ID denotes a job that contains manually validated peptide matches. By clicking on an ID link, the Proteins Overview page opens.

Note that Mascot, SEQUEST-Bioworks or X!Tandem output can be compared to your Phenyx results. The data must first be converted into a Phenyx job using the Import tool on the Management Console page. Click on the Import link under the Jobs Management menu. Select the relevant protein identification software in the Import from drop-down list. If working with Results data, then click on the Browse button to locate the Local File or archive on your hard disk. Otherwise, copy and paste the Remote URL of a Mascot Peptide Summary Report, a SEQUEST HTML Summary File or a SEQUEST-Bioworks XML Summary file in the Add URL box. If your Mascot or SEQUEST-Bioworks data are only accessible using a login and password, then type http://login:password@result_url instead of a basic http://result_url. Click on the Import button to convert the file into a job with a new identification (ID) number. Go to the Phenyx Desktop to start working with the job. In the Title column of the Job Report, an information about the origin of the imported job is given, for example the original ID of an imported Phenyx job is referenced [in brackets]. In case of an imported Mascot job, the Title column displays [MASCOT IMPORT] and file title. In the Comment column of the Job Report, information about the imported file or URL is given.

Remove jobs from a comparison table

To withdraw a particular job from an existing table, deselect (uncheck) the ID number in the Jobs List and click the Update button. The job(s) disappear from the table but not from the Jobs List. Note that this action also directly effects the Detailed Results Comparison page.

Empty List: Clears the Jobs List and table in order to start again.

Results comparison table

Identified proteins for the different jobs are compared in the table. By default, the results comparison table is sorted according to Group (smallest to largest number). This sort order can be reversed (largest to smallest number) by clicking on the header of the Group column. Otherwise, sort the rows in ascending order (A to Z, 0 to 9) or descending order (Z to A, 9 to 0) by clicking (or re-clicking) on the header of any column.

Protein AC: Database accession number. By clicking on the AC link, the Detailed Results Comparison page opens. The suffix _WOSIG0 denotes the protein sequence without the Signal sequence. The suffix _WOPP0 denotes the Peptide 1 in SwissProt. The suffixes _CHAIN0 or _VAR1 denote Chain 1 from the original protein entry or Variant 2 from the original protein entry, respectively.

To change the information displayed in the Protein AC column, go to the Selected Proteins menu on the left side of the Results Comparison page. Select one of the following choices and click the Update button. Deselect an item and click the Update button to remove it from the table.

- All: Every protein identified in each job is listed in the table.
- Main Proteins: Only the proteins with the best scores for each job are displayed (=Main protein of the Proteins Overview page).
- Proteins in all Jobs: Only the proteins present in all the selected jobs are shown in the table.
- Show Invalid Proteins: Check the box to display proteins that were invalidated automatically by Phenyx or manually by the user.

Group: A grouping is assigned when jobs share a common AC and/or have the same peptide matches. The Group with the smallest number has the largest number of common denominators between the compared jobs.

Description: Official protein name followed by common synonyms (in parentheses).

In the job columns, the number of valid peptides is given by default. By clicking on the number link, the Protein Match Details page opens. White

boxes display information from subset proteins, the same sub-entries listed under Subset in the right pane of the Proteins Overview page. By contrast, blue boxes label main proteins, the same listed in the best-scoring protein table of the Proteins Overview page. To change the information displayed in the job column, go to the Display Options menu on the left side of the Results Comparison page. Select (check) one or more of the following choices and click the Update button. If more than one item is listed, then they are separated by hyphens (/) in the boxes of the table. Deselect (uncheck) an item in the menu and click the Update button to remove it from the table.

- # Valid Peptides: Number of valid peptide matches found for the given protein.
- # Peptides: Total number of peptide matches found for the given protein.
- Protein Score: Effective score, the protein score that is recalculated based on user-validated peptides. It is the sum of the best scores for validated peptide sequences.

Protein detail list

A supplemental view of the comparison according to peptide matches can be obtained on the Detailed Results Comparison page. Select (check) the protein(s) to be expanded upon in the Protein AC column. The AC number is listed under the Protein Detail List on the left side of the page. Deselect (uncheck) the proteins in the Protein AC column to remove them from the list. Click the Add button to see the elaboration. The first time this action is performed, the selections are stored in memory. Therefore, more proteins can later be added to the original Detailed Results Comparison page by making selections and then clicking the Add button. This buffer can only be cleared by clicking the Empty List button on the Detailed Results Comparison page.

A different, additional Detailed Results Comparison page can be viewed by selecting or deselecting proteins in the Protein AC column. Click the Open button to display the listed proteins in a new window.

Export options

Possibilities are offered to transfer the Phenyx comparison data to another software platform for further analysis or publication purposes.

CSV: Comma Separated Value format. Click the CSV link to open or save the table as a delimited text file that uses a comma (,) to separate the data.

Excel: Exports the results of the comparison as an Excel (.xls) spreadsheet in either the Microsoft application or a new browser window. Use the Save As option to store a copy of the file to your hard disk

results comparison

Detailed Results Comparison page

Updated: May 10, 2008

This page allows the user to view one or more job results in tabular format. The table is organized according to identified peptide matches for those proteins and jobs currently selected on the Results Comparison page. It is good habit to use the refresh function in your browser window in order to have any changes made to the selections on the Results Comparison page be reflected on the Detailed Results Comparison page.

Detailed results comparison table

The peptide matches for selected proteins and jobs are compared in the table. Listed peptides in the Sequence column are the reference for this table. Protein AC is the default choice in the Group by menu on the left side of the page. The table is then arranged by protein accession number.

Each job that was active in the table on the Results Comparison page is assigned an integer (for the sake of clarity). The Jobs List on the left side of the page is simply a legend explaining the representations. The Proteins Overview page can be directly accessed by clicking on the ID links in the Jobs List. In the job columns of the table, the validity (plus sign, +) or invalidity (minus sign, -) of the match(es) is given. By clicking on the + or - link, the Peptide Match Details page opens.

Alternatively, the table can be arranged by job. Go to the Group by menu on the left side of the Detailed Results Comparison page. Select Job ID and click the Update button. The validity (plus sign, +) or invalidity (minus sign, -) of the match(es) is now given in the Protein AC columns. Select Protein AC under the Group by menu and click the Update button in order to revert back to the default view.

To change the information displayed in the table, go to the Display Options menu on the left side of the Detailed Results Comparison page.

Select one of the following choices and click the Update button. Deselect an item and click the Update button to change the view of the table.

- Display AA Mods: Bold modified amino acids in the Sequence column.
- z-Score: The z-Score is reported in the job columns. By clicking on the z-Score link, the Peptide Match Details page opens.
- Best Peptide Match: Show only the best match for each peptide.

All of the AC numbers selected to appear in the table are shown in the Protein AC List found at the left side of the Detailed Results Comparison page.

To delete a particular protein from the existing table, deselect (uncheck) the AC number in the Protein AC List and click the Update button. The protein(s) disappear from the table and the Protein AC List.

Empty List: Clears the Protein AC List and table in order to start again.

Export options

Possibilities are offered to transfer the Phenyx comparison data to another software platform for further analysis or publication purposes.

CSV: Comma Separated Value format. Click the CSV link to open or save the table as a delimited text file that uses a comma (,) to separate the data.

Excel: Exports the results of the comparison as an Excel (.xls) spreadsheet in either the Microsoft application or a new browser window. Use the Save As option to store a copy of the file to your hard.

Understanding Phenyx: the scoring system

Updated: May 10, 2008

This section is intended to assist you in gaining a greater facility and knowledge of Phenyx. It contains more in-depth explanations on important topics of interest. It also encourages you to think critically about your job submissions and results.

The MS/MS scoring model

Journal papers and conference proceedings fully describing the MS/MS scoring model used in Phenyx are available to the public. The references can be found at http://www.phenyx-ms.com/documentation/documentation.html.

In brief, Phenyx computes a score to evaluate the quality of a match between a theoretical and experimental peak list (i.e. mass spectrum). A match is thus a collection of observations deduced from this comparison. The basic peptide score is ultimately transformed into a normalized z-Score and a p-Value.

A basic peptide score is the sum of raw scores for up to twelve physico-chemical properties such as:

- Presence of specific fragment ion types (a, b, y, y++, b-H2O, etc.)
- Co-occurence of ion series (modeled by Hidden Markov Models)
- Relative intensity versus the fragmentation type
- Amino acid modifications (PTMs, chemical modifications, etc.)
- Peptide amino acid composition
- Peptide and fragment mass error
- Number of missed cleavages

Each of these properties are computed during the comparison between theoretical and experimental data. The influence of each component depends on the parent ion charge. Some of the properties are parameters that the user can define on the Submission page.

How a score is computed

A score decides whether a match is correct or not. The relative influence of each of the components is difficult to decide *ex nihilo*. But Phenyx measures the probabilities of observing correct (H1) and wrong (H0) matches and describes the score as a log likelihood ratio (Score L):

$$\text{Score L} = \log \frac{P(M,s,H1)}{P(M,s,H0)}$$

where P(M,s,H1) is the probability to observe the match M under the hypothesis H1 that the peptide s is correct, and P(M,s,H0) is the probability under the null hypothesis H0 that the peptide s is random.

How a score is normalized

The scores are normalized in order to be able to compare matches under various experimental conditions. For each peak list Phenyx computes a score distribution for a search in a given database. Then it computes a score distribution on a randomly sampled set of peptides to provide a random distribution. Finally it normalizes the original scores to the random distribution (Figure 1). The normalized peptide score is called the z-Score.

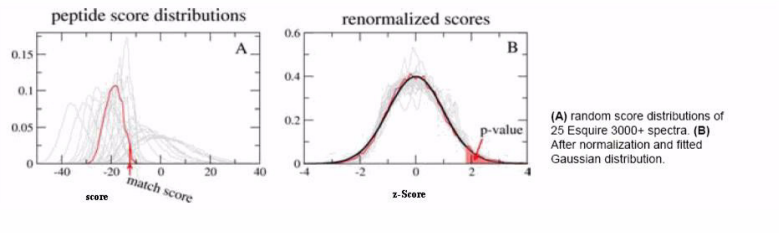


FIGURE 1. Graph A represents the random score distributions for twenty-five Esquire 3000+ spectra. Graph B is the product of normalization and fitting to a Gaussian distribution.

The z-Score therefore represents the distance from a random match. However, the normalized score is not always sufficient enough to assess the correctness of a match because the accuracy is dependent on the absolute size of the search space. As a probabilistic algorithm, Phenyx therefore associates a p-Value to each match.The p-Value is the probability to obtain a score greater than or equal to a given score by chance. The p-Value is approximately $1-(1-p)^N$ where p is the probability to get a given score as random and N is the number of peptides considered in the population to be matched. The lower the p-Value, the more significant the match.

The p-Value is influenced by the size of the search space. A match is unlikely to be random for a small set of sequences. For a given z-Score, a much better p-Value is retrieved (with a low value) in the Swiss-Prot database (two hundred thousand entries) than in the NCBI database (several million entries).

How Phenyx retrieves amino acid modifications not specified by the user

Phenyx takes full advantage of the annotated sequence information found in the Swiss-Prot and TrEMBL databases such as post-translational modifications (PTMs), binding sites, variants, alternative splicing of variants, etc. During a search, Phenyx looks at the difference between an experimental peptide mass and a theoretical peptide mass in order to determine modifications to the protein sequence. If a mass difference corresponds to a known PTM that is annotated in Swiss-Prot (even if the modification was not selected by the user on the Submission page), then the peptide sequence is considered modified and reported in the results. These modifications are always treated as variable by the scoring model.

An example of this functionality is shown in Figure 2. The MS/MS data were collected from a human blood sample that had undergone proteolytic digestion. The set of submission parameters used to produce the data is given in the table below.

Parameter	**Setting**
Database	Swiss-Prot

Parameter	**Setting**
Instrument Type	ESI Ion Trap
Scoring Model	HCTultra
Default Parent Charge	1,2,3
Trust Parent Charge	Yes
Number of Rounds	1
Variable Modifications	Oxidation_M Oxidation_HW
Enzyme	Trypsin_(KR_noP)
Missed Cleavage(s)	1
Cleavage Mode	Normal
Turbo	Tolerance=800.0 ppm Coverage=20% Series=b;b++;y;y++
Conflict Resolution	Yes
Parent Error Tolerance	2.0 Da
Min. Peptide Length	6
Min. Peptide z-Score	5.0
Max. Peptide p-Value	1.0E-6
AC Score	7.0

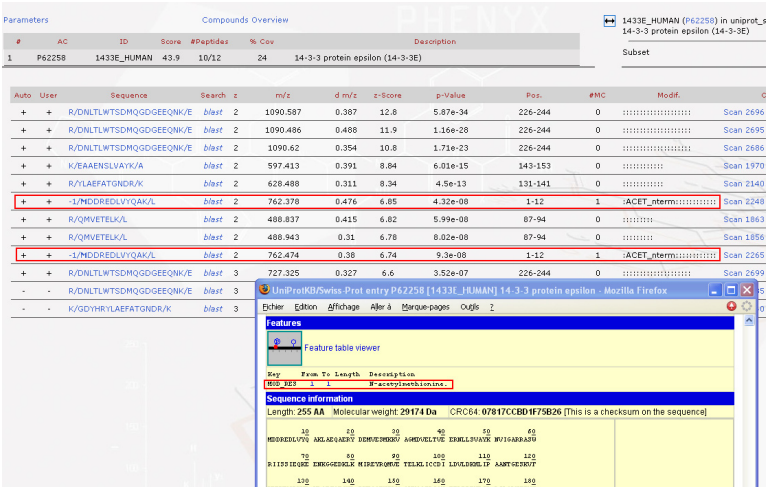


FIGURE 2. Although not specified on the Submission page, the N-terminal modification of methionine (M) is retrieved by Phenyx. The Swiss-Prot protein entry is inset in order to show the modified residue information found in the database's Feature Table (FT). The acetylated peptides (enclosed in red) are found in the Modif. column at the bottom of the Proteins Overview page.

Reproducibility of z-Scores and p-Values

The random peptide population generated during scoring to determine the probability of a random match against a random database is unique for each calculation. Consequently, a small variation will occur in the results when resubmitting a job using the same parameters and peak lists. This variability does not usually exceed ten percent of the z-Score value. However, variability increases when searching a restricted set of

proteins such as one accession (AC) number or a narrow taxonomy range (Figure 3).

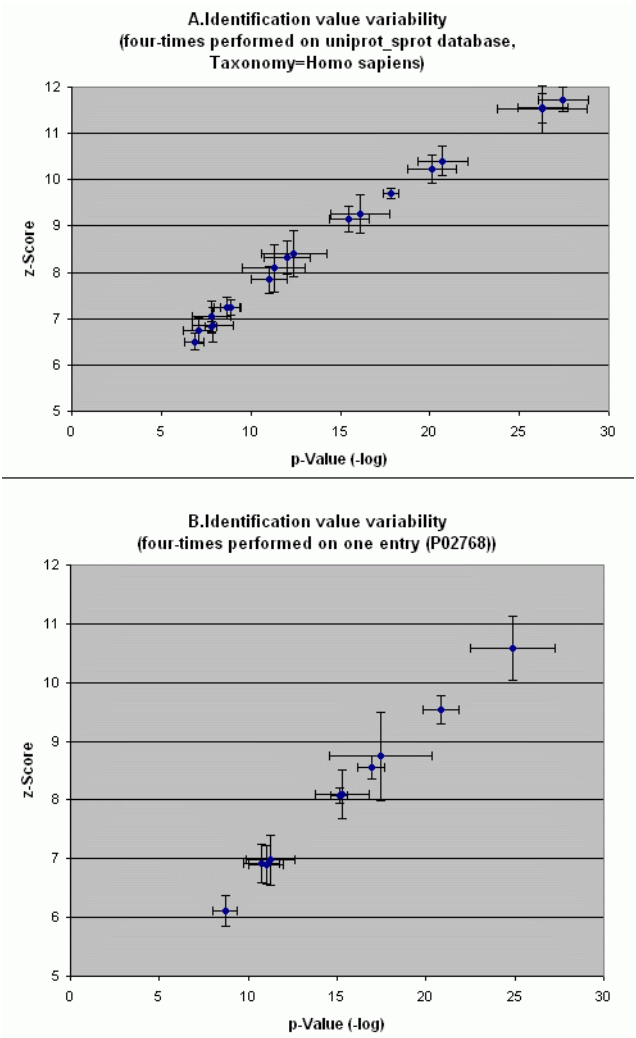


FIGURE 3. Searches A and B were submitted four consecutive times. Each point in the graphs represents the average z-Score and p-Value for one compound (i.e. a single peptide identification). The variability is expressed by the error bars

(based on standard deviation) displayed for both the x and y values. All human protein entries in the Swiss-Prot database were searched in A. Only one protein entry was searched in B, so the number of potential peptide sequences is relatively small. Phenyx introduced random peptides to the set of theoretically digested peptides for this protein in order to reach an adequate number of peptides to perform scoring on. The reproducibility of the results is observed to be dependent on the size of the sampling set; worsening for limited searches.

Available scorings

The type of instrument used to acquire the data determines the choice of the appropriate scoring to perform the search. Each scoring takes into account a specific set of ion fragment series. The fragment error tolerance is already included in the selected algorithm.

Scoring description as displayed into the Scoring Model top-down menu refers to the fragmentation process. Where the parent mass is detected is not considered. For instance, the scoring model CID_LTQ_scan_Orbitrap_6ppm refers to data that have been fragmented in LTQ and detected in Orbitrap with a fragment error tolerance as up as 6ppm.

Phenyx scoring list is in progress and regularly updated. Please contact us (support@phenyx-ms.com) if your type of instrument is missing.

Understanding Phenyx: the effect of submission parameters on result accuracy

Updated: May 10, 2008

The general expectation from Phenyx is that the identified proteins and peptides accurately describe your experimental sample. There are unstable factors that impact the accuracy such as the quality of the mass spectra and databases used in a search. It is also known that the specification of certain parameters by the user on the Submission page can influence the correctness of a search. These submission parameters are described below in order to help you make informed decisions. When evaluating the results of different searches, the z-Score and/or the p-Value are the best standards for separating correct matches (true positives) from random matches (false positives).

Limit the search space size

The z-Score is a normalized score and thus independent of the size of the search space. It mainly depends on the random matches because the basic peptide score is normalized to a random distribution. In contrast, the p-Value is dependent on the search space size. The p-Value deteriorates as the search space grows larger because the chance of a peptide being a random match increases.

The size of a search space can be defined by the following parameters:

- Taxonomy and/or AC List: The more specific the selection, the smaller the database size.
- Cleavage Mode: A half cleaved selection generates more theoretical peptides to be sampled, thus expanding the search space.
- Number of Missed Cleavages: Increasing the number of missed cleavages increases the number of peptides to be considered.

- Number of Variable Modifications: The greater the number of modifications to look for, the larger the sample set of theoretical peptides.

Examples of the effects of each of these parameters follow (Figure 1 to Figure 4). The MS/MS data were collected from a human blood sample that had undergone proteolytic digestion. The set of submission parameters used to produce the data is given in the table below.

Parameter	**Setting**
Database	Swiss-Prot
Instrument Type	ESI QTOF
Scoring Model	Default
Default Parent Charge	2,3
Trust Parent Charge	Medium
Number of Rounds	1
Enzyme	Trypsin_(KR_noP)
Missed Cleavage(s)	1
Turbo	Tolerance=800 ppm Coverage=20% Series=b;b++;y;y++
Conflict Resolution	Yes
Parent Error Tolerance	2.5 Da
Min. Peptide Length	6
Min. Peptide z-Score	5.0
Max. Peptide p-Value	1.0E-6
AC Score	6.0

Note that the scale of the z-Score versus p-Value plots are not equal in each of the examples. The various parameters do not have the same impact on the curve fitting.

Change the taxonomy or AC list

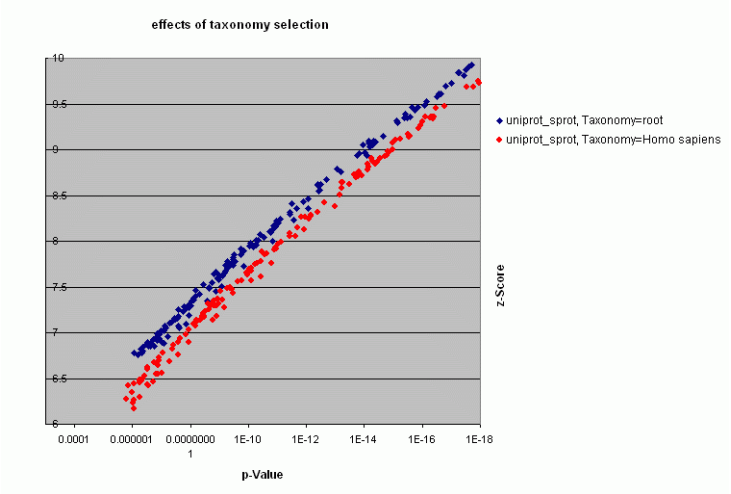


FIGURE 1. This graph shows the effect of taxonomy on pairs of z-Scores and p-Values. The Root selection means that the full Swiss-Prot database was searched. Selecting Homo sapiens reduced the size of the database by a factor of about 13. The curve slowly moves to the right part of the graph, meaning the smaller the size of the search space the lower the p-Value for a given z-Score.

Figure 1 demonstrates how reducing the search space size by a taxonomy restriction can generate better p-Values. In other words, a match is of higher quality when there is less chance of being random.

Change the cleavage mode

Three searches were performed in order to produce the next graph. In addition to the settings described above, the following parameters were tweaked:

Parameter	**Job 1**	**Job 2**	**Job 3**
Taxonomy or AC List	Homo sapiens	Homo sapiens	P02768
Cleavage Mode	Half cleaved	Normal	Normal

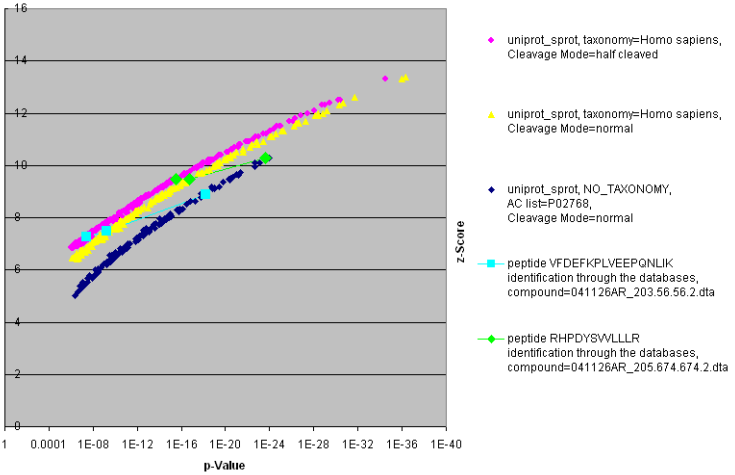


FIGURE 2. Peptide identifications show that the curve climbs toward the right part of the graph as the search size decreases. For a given z-Score, the p-Value gets better (becomes smaller) as the search size decreases. The z-Score/p-Value pairs for two peptide matches belonging to serum albumin (P02768) are highlighted for the three different searches.

The cleavage mode has one of the most significant impacts on the results. The number of identified peptides can improve, which in turn has favorable repercussions on the protein coverage and score (if the peptide matches are valid). However, by going from normal to half cleaved, the number of theoretical peptides generated can increase by a factor of ten. The increased search space size causes the p-Values to deteriorate (i.e. become larger). The combined use of cleavage mode and taxonomy can counterbalance these effects. For example, a restricted taxonomy selection will limit the search space size and ameliorate the p-Values while searching in half cleaved mode.

In Figure 2, the p-Value decreases from Job 1 to Job 3 as the number of searched peptides decreases. The z-Score remains relatively constant when searching many peptides (Jobs 1 and 2) but deviates slightly when only searching one protein entry (Job 3). This deviation is caused by undersampling during the random process used to calculate the z-

Scores. When searching only one accession number, it is recommended to concentrate on the interpretation of the p-Values.

Change the number of missed cleavages

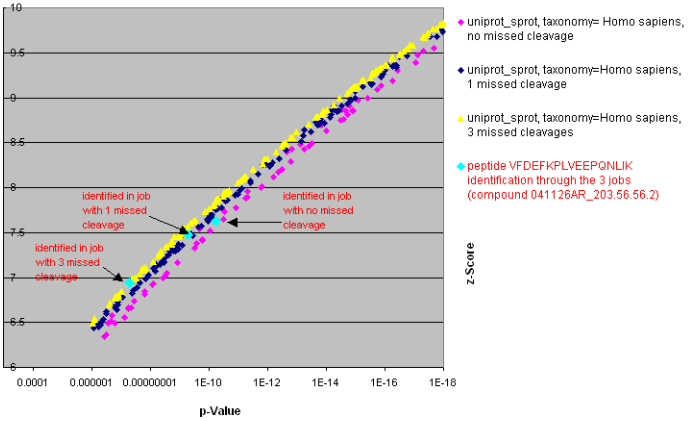
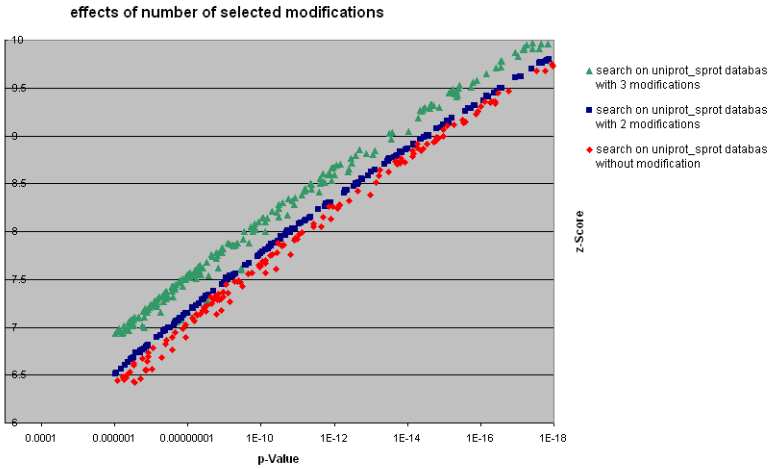


FIGURE 3. Going from 0 missed cleavages to 1 missed cleavage multiplies the probability of false positives by about two, as the search space grows by a factor of about two. In other words, as the number of missed cleavages increases, the worse the p-Value for a given z-Score. Note that the peptide VFDEFKPLVEEPQNLIK is identified when zero missed cleavages are allowed since the rule for trypsin states that no cleavage occurs if proline (P) is C-terminal to lysine (K) or arginine (R) (this was fixed by the regular expression ^P on the Management Console page).

The higher the number of missed cleavages selected, the larger the sample set of theoretical peptides. This introduces a more important random component into your search process. As seen in Figure 3 the p-Value becomes worse for a given z-Score. The best practice is to use up to one missed cleavage, allowing the search to truly reflect enzyme reactivity while working within the theoretical confines of Phenyx. The effect of the number of missed cleavages on result accuracy is however less than the effects of the selection of taxonomy or cleavage mode.

Change the number of variable modifications

Three searches were performed in order to produce the next graph. In addition to the settings described above, the following parameters were tweaked:



Parameter	**Job 1**	**Job 2**	**Job 3**
Variable Modifications	Oxidation_M PHOS Cys_CAM	Oxidation_M Cys_CAM	None

FIGURE 4. The number of variable modifications increases the set of sampled peptides, thus increasing the size of the search space. As the number of theoretically modified peptides increases, the more up and left the curve is displaced. In other words, the p-Value worsens for a given z-Score.

The increase in search space from Job 2 to Job 3 (i.e. searching for the oxidation of methionine and the carboxy methylation of cysteine) is about a factor of two to three which has a relatively small impact on the p-Values. However, when a variable modification is added such as phosphorylation that looks at the less common amino acids serine (S), threonine (T) and tyrosine (Y), the search space increases significantly.

The overall level of confidence in matches decreases, even for the unmodified peptides.

Search in one round or two

Phenyx performs searches in one or two serial steps called rounds. A classical one-round search is performed on a defined set of proteins (i.e. when a full database or taxonomy or accession number (AC) is selected). Identified peptides that fulfill the z-Score and p-Value thresholds are accepted and the corresponding protein is validated.

The basic principle of two rounds is that the first round processes all the proteins in the designated search space and the second round only processes the proteins that passed the first round. The first round parameters need to be stringent enough to sufficiently validate protein identification (i.e. a few good peptides). The second round parameters enable you to open the search criteria in order to increase the sequence coverage, by searching for combinatorial modifications or other special features. A two-round search therefore identifies proteins according to a first set of parameters and then performs a more exhaustive search on the proteins while saving computation time and reducing the random match rate.

One-round usage

A single round is adapted to the following cases:

1. You submit a large set of data and your aim is simply to retrieve a list of identified proteins.

The best strategy is to use a set of search parameters that will minimally validate your data. Your goal is not to optimize the sequence coverage, but to identify a protein with peptides having at least a certain level of confidence.

2. You wish to identify proteins from a low complexity sample (e.g. a 2D spot).

Broaden your query by searching large databases. You should select parameters that are not too stringent assuming that you will have to manually check the results to retrieve putative proteins.

3. You wish to retrieve one or two known proteins (i.e. you are able to specify the AC List for two proteins).

Since you know which proteins you are looking for, you can specify the search parameters in a single round. The lower the acceptance criteria, the greater the peptide coverage. The number of manual peptide validations needed to be done will also be higher in order to reduce the false positive rate.

Two-round usage

Two rounds are adapted to the following cases:

1. You have a restricted sample (e.g. a 2D spot) and wish to raise the level of confidence in your results through increased sequence coverage.

You should select stringent parameters in the first round and apply looser parameters in the second round to fine-tune your results. Your aim is to find out what peptides can be generated from your protein mixture.

2. You are looking for special or difficult peptides in your mixture.

You should be able to retrieve peptides with high combinatorial properties (i.e. several variable modifications) or peptides that resulted from nonspecific cleavage (half cleaved). The first round searches for obligatory peptides such as unmodified, ideally cleaved peptides or isotopically labeled peptides from quantification studies. The second round looks to see if the identified proteins can be covered by other identified peptides.

Remember that the two-round process only validates proteins found in the first round. If your first round criteria are too tolerant, the false positive rate will increase since the second round adds value to the proteins retrieved in the first round.

The appropriate use of a two-round search should reduce the overall calculation time. Specify stringent criteria to search in the first round and only query a small subset of a database with computation intense settings in the second round.

Two-round search parameters

The most important parameters are listed below as a guideline for searches done on data from a classical protein identification prospective (e.g. MudPIT or a gel-based workflow) or for the identification of a protein with minimal criteria to optimize sequence coverage.

- Fixed Modification(s): If you specify a fixed modification in the first round, it must also appear in the second round.
- Variable Modification(s): It is generally better to keep variable modification(s) for the second round. You will insure the retrieval of modified peptides for proteins accepted in the first round, without increasing the search space size (i.e. without introducing a higher random match rate or longer calculation times). The oxidation of methionine is an exception. This modification does not significantly impact the calculation time and might be contained in some single hit identifications.
- Missed Cleavage(s): It is advised to limit the number of missed cleavages to 0 or 1 in the first round. More missed cleavages are better suited to a second round because they are often not considered determinant in the identification of a protein.
- Cleavage Mode: It is best to use the Half cleaved option in the second round for the same reasons as Missed Cleavages. Proteins are usually validated with peptides cleaved strictly according to the rules. A protein is rarely validated based on nonspecifically cleaved peptides alone.
- Conflict Resolution: This feature should be selected in the second round. Since solving conflicts reduces the number of peptides, it is more appropriate to use it at the end of a search to avoid reducing the number of proteins that pass the first round.
- Peptide z-Score and p-Value: Looser parameters can be set in the second round. Since a protein can be validated by a few good peptides, it is more acceptable to associate peptides with lower scores to an already validated protein. As the z-Score threshold decreases, the inclusion of peptides in a result and the p-Value (taken together with the outcome of conflict resolution) determine which peptides contribute to a protein score. You can adapt these values to your own validation rules.
- Turbo: As a filtering factor, this feature should be used in the first round. While it rejects the interpretation of spectra with poor

matches, it accelerates the first round. The only drawback is that it may fail to propose interpretations for low quality spectra.

Activate conflict resolution

Conflicts arise when the scoring algorithm can match more than one peptide with acceptable z-Score and p-Value to a given spectrum. The available data (m/z, intensity signal or scoring model) is not specific enough to generate a unique peptide identification. There are many reasons why a mass spectrum may be represented by multiple peptide sequences originating from different proteins. Some of these reasons are:

- Peptides which interchange isobaric amino acids such as leucine (L) and isoleucine (I) or without sufficient mass-to-charge resolution to distinguish lysine (K) and glutamine (Q).
- Sequences in which amino acid compositions are identical but arranged in a different order and there is not enough signal present in the spectrum to discriminate them.
- Peptides with similar m/z and scores which derive from diverse sequences. They share some of the same peaks. This happens for spectra that cover only partial sequences.

When faced with an unresolved conflict, you are the most qualified to make sound judgments based on your experimental experience, understanding of the concepts and level of confidence in the selected databases.

How Phenyx resolves conflicts

When the conflict resolution functionality is activated on the Submission page:

1. Resolving a conflict is considered only if the conflicting peptide matches are of good enough quality, i.e. if they reach a minimum z-Score and p-Value. These thresholds are a function of the parent charge and are set by Phenyx. If the peptides' z-Scores and p-Values are too low, then the conflict is not resolved and the matches get a rejected status. The level of confidence is therefore not considered as high enough.

2. When peptide matches have sufficient scores to be resolved, a discrimination rule is applied based on a z-Score difference and a p-Value ratio. These values are a function of parent ion charge. They are configurable values that can be set by a Phenyx administrator. If the difference between the highest scoring peptide and the compared one(s) is larger than the discrimination values, then the conflict is considered as solved. A winner and a loser are defined. A winner will keep its valid status (therefore contributing to the protein score) and a loser will get an invalid status (and will not contribute to the protein score).

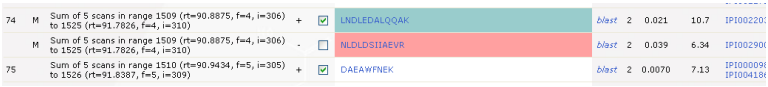

74 M Sum of 5 scans in range 1509 (rt=90.8875, f=4, i=306) to 1525 (rt=91.7826, f=4, i=310) + LNDLEDALQQAK blast 2 0.021 10.7
M Sum of 5 scans in range 1509 (rt=90.8875, f=4, i=306) to 1525 (rt=91.7826, f=4, i=310) - NLDLDSIIAEVR blast 2 0.039 6.34
75 Sum of 5 scans in range 1510 (rt=90.9434, f=5, i=305) to 1526 (rt=91.8387, f=5, i=309) + DAEAWFNEK blast 2 0.0070 7.13

FIGURE 5. On this Compounds Overview page there is one resolved conflict for Compounds 74. The z-Scores and p-Values for both peptides were good enough to apply the conflict resolution algorithm. Phenyx then evaluated the quality of each conflicting interpretation in order to find out which match could be considered as correct. In both examples, the z-Score difference is large enough to resolve the conflict. The winner sequence is highlighted in green and has a valid (+) status. The loser sequence is highlighted in red and has an invalid (-) status.

3. When the scores of conflicting peptides (with *a priori* valid status and scores above the Phenyx thresholds) are very similar, but the score difference or p-Value ratio is below the discrimination threshold Phenyx decides that it cannot unambiguously resolve the conflict. The peptides are considered as multiple winners. They are labeled as conflict winners and can be subjected to manual validation for a final decision.

87 Sum of 10 scans in range 2012 (rt=118.362, f=4, i=413) to 2053 (rt=120.624, f=5, i=417) + IIAATIENAQPILQIDNAR blast 2 0.013 7.75
88 Sum of 10 scans in range 2014 (rt=118.474, f=2, i=418) to 2054 (rt=120.68, f=2, i=427) + IIAATIENAQPILQIDNAR blast 3 0.016 9.59
89 M Sum of 12 scans in range 2131 (rt=124.86, f=3, i=607) to 2174 (rt=127.234, f=2, i=452) + LALDLEIATYR blast 2 0.011 9.21
M Sum of 12 scans in range 2131 (rt=124.86, f=3, i=607) to 2174 (rt=127.234, f=2, i=452) + LALDIEIATYR blast 2 0.011 9.21

FIGURE 6. On this Compounds Overview page there are two conflicting sequences for Compound 89, both highlighted as winner sequences. The z-Scores and p-Values for both sequences were good enough to apply the conflict resolution algorithm. However, the z-Score difference is not large enough to resolve the conflict. Both sequences are considered as winners and highlighted in green. Note that the only difference in these particular sequences is an isobaric leucine/isoleucine substitution.

Some points to consider

There are several generalities that you should bear in mind:

- A protein must have at least one valid match and satisfy the threshold AC Score (protein score, specified by the user on the Submission page) in order to be listed as a result with all of its potential peptide matches.
- Only valid matches and conflict winners are used to calculate the protein score.
- The Phenyx threshold values for conflict resolution might be set higher than z-Score and p-Value thresholds specified by the user on the Submission page. Therefore it is possible that a couple of peptide matches will lose their valid status when the Conflict Resolution function is activated.
- A conflict loser will never contribute to the respective protein score (AC Score) even if it has acceptable submission z-Score and p-Value criteria. A protein can therefore become invalidated since its remaining peptide matches might not be sufficient to attain the minimum AC Score specified by the user on the Submission page.
- A rejected match can also be in conflict with another peptide(s). The definition of rejected takes precedence over conflict rulings.
- In queries performed using two rounds, it is recommended to activate Conflict Resolution in the second round only. Otherwise, you run the risk of obtaining a disproportionate number of rejected matches or of losing protein identifications that do not have at least one valid match for lower quality spectra.

Index

A

AA modif. details 13
AC 39, 45, 52
AC list 10, 74
AC score 18
Acceptance parameters 18
Account settings 32
Actions on selected 59
Actions on selected job(s) 4
Add 62
Add following jobs 59
All 4, 61
Amino acid modifications 13
Annotated amino acid modifications 69
Apply changes 56
Auto 41, 46, 51

B

Best Peptide Match 66
Best-scoring protein table 39
BLAST 42, 46, 52
Browse files 32

C

Change password 32
Charge state 57
Cleavage mode 16, 74, 82
Cleavage site 23
Cleavages enzymes 23
Comment 5
Compare 4, 59
Completed 4, 5
Compound 44, 51
Compound description 55
Compounds overview 6, 39
Conflict resolution 17, 82, 83
Convert MS/MS peak lists 35
Coverage 40
CSV 63, 66

D

Databanks 32
Databases 8
Date 5
DB 45
Def. 13, 14
Default parent charge 11
Delete 4
Delete profile 8
Delta m/z 42, 46, 52, 56
Delta mass 57

Description....40, 45, 61
Deselect all....4, 53
Detailed results comparison....62
Detailed results comparison table....65
Discrimination rule....84
Display AA mods....66
Display following AC....53
Display options....65
Documentation....3
DoNotCleave enzyme....16
Down....4

E

Empty list....60, 66
Enzyme....14
Enzyme definitions....14
Enzyme inefficiency....16
Error....4, 5
Excel....63, 66
Experimental charge....55
Experimental m/z....55
Export....63, 66
Export Excel....6
Export files....33
Export last saved validated peptides....53
Export text....6
Export XML....6

F

False positive rate....33
FASTA format....8, 32
File format....19
Fixed modifications....13, 82
Fragment ion table....56

G

Gain....24
Group....61
Group by....65
Group defs....32

H

Hide user invalidated entries....53

I

ID....5, 39, 45
Immonium ions....58
Import....34, 60
Insert jobs in a results comparison table....59
Instrument type....11
Ion series selection....56
IPI....9
Isobaric amino acids....83
iTRAQ....7

J

Job ID 65
Job investigator 35
Job menu 6
Job report 3
Jobs list 60
Jobs management 32, 60

K

Kill 4

L

Log out 3
Logged as 3

M

m/z 42
Main proteins 61
Management console 3
Mascot 60
Mass spectrum 56
Match score 56
Max intensity 56
Maximum peptide p-Value 18, 82
Minimum peptide length 18
Minimum peptide z-Score 18, 82
Missed cleavages 16, 74, 82
Modification definitions 13
Modification name 13
Modifications 43, 47, 55
MS/MS scoring model 67
Multi 51
My defs 23

N

Name 23
Navigation bar 3
NCBInr 10
Neutral losses 58
Normalized score 68
Number of missed cleavages 42, 47
Number of peaks 55
Number of peptides 40, 62
Number of rounds 12, 80
Number of valid peptides 62

O

One round 80
Open 62

P

Parameters 4, 6, 38, 39
Parent error tolerance 18

Partial cleavage 16
Peak intensity 57
Peak lists 19
Peak mass 57
Pending 4, 5
Peptide match table 41, 46
Peptide score 67
Position 42, 47
Precursor masses 58
Private databank 32
Profiles 8
Protein AC 61, 65
Protein AC list 66
Protein detail list 62
Protein details 40
Protein information table 45
Protein score 62
Proteins in all jobs 61
Proteins overview 6
PTMs 13
p-Value 42, 47, 69

R

Refresh 4
Remove jobs from a results comparison table 60
Reproducibility 71
Reset to auto validation 53
Reset zoom 56
Residue modifications 27
Resubmit 6
Resubmitted peak lists 19
Results comparison table 61
Round 1 12
Round 2 12
Running 4, 5

S

Save 53
Save as profile 8
Score 40, 46
Score calculation 68
Score normalization 68
Scoring model 11, 67, 73
Scoring system 67
Search 42, 46, 52
Search engine 11
Search space size 74
Select all 53
Sequence 41, 46, 52, 55, 65
Sequence views 48
SEQUEST 60
Server log 19
Set as default 8
Show invalid proteins 61
Status 5
Submission 3

Submit 19
Subset 40
Suffix 45
Swiss-Prot 9
Swiss-Prot feature table 69
Synonyms 45

T

Taxonomy 10, 74
Theoretical charge 56
Theoretical mass 55, 57
Title 5, 8
Tolerance 14
Top 4
TrEMBL 9
Trust parent charge 11
Turbo 17, 82
Two rounds 81
Type 13
Type of fragment 57

U

UniProt 9
UniProt_Swiss-Prot, reversed 9
unmatched fragment spectra list 34
Up 4
Update 59
User 5, 41, 46, 51
User Permissions 7
Utilities 35

V

Validate peptide matches 54
Validity 41, 46, 51
Value types 57
Variability 71
Variable modifications 13, 75, 82

X

X!Tandem 60
XML total def 30
XML user's def 30

Z

z 42, 46, 52
Zoom factor 56
z-Score 18, 42, 52, 66, 68